43561

National Library of Canada
Canadian Theses Division
Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada
Division des thèses canadiennes

PERMISSION TO MICROFILM — AUTORISATION DE MICROFILMER

- Please print or type — Écrire en lettres moulées ou dactylographier

Full Name of Author — Nom complet de l'auteur
Dale Christian Stevenson

Date of Birth — Date de naissance
2 July 1945

Country of Birth — Lieu de naissance
Canada

Permanent Address — Résidence fixe
515 rue Francois, #108, Nuns Island, Verdun, P.Q. H3E 1G5

Title of Thesis — Titre de la thèse
Categorical Perception and Selective Adaptation Phenomena in Speech

University — Université
University of Alberta, Edmonton, Alberta, Canada

Degree for which thesis was presented — Grade pour lequel cette thèse fut présentée
Doctor of Philosophy

Year this degree conferred — Année d'obtention de ce grade
1979

Name of Supervisor — Nom du directeur de thèse
Dr. A.J. Rozsypal

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

Date
24 October 1979

Signature
Dale C Stevenson

NL-91 (4/77)

National Library of Canada

Cataloguing Branch
Canadian Theses Division

Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada

Direction du catalogage
Division des thèses canadiennes

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

NL-339 (3/77)

THE UNIVERSITY OF ALBERTA

CATEGORICAL PERCEPTION AND SELECTIVE ADAPTATION PHENOMENA IN SPEECH

by

DALE C. STEVENSON

©

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

SPEECH PRODUCTION AND PERCEPTION

DEPARTMENT OF LINGUISTICS

EDMONTON, ALBERTA

FALL, 1979

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled "Categorical Perception and Selective Adaptation Phenomena in Speech" submitted by Dale Christian Stevenson in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Speech Production and Perception.

A Rozsypal

Supervisor

John T. Hogan

R. E. Rink

External Examiner

DATE June 26, 1979

ABSTRACT

A model of categorical perception is presented in which the mapping between a physical acoustic continuum and the corresponding perceptual continuum is assumed to be nonlinear. It is proposed that such nonlinearity is sufficient to account for many, although not necessarily all, cases of bserved categrical perception. A new experimental paradigm is described for testing this model of categorical perception. Stimuli are constructed by adding together the waveforms of two speech signals with relative weights α and β such that the composite waveform s' is described by $s' = \alpha s_1 + \beta s_2$, where s_1 and s_2 are the formant transitions from two initial or final stop consonants, and $\beta = 1-\alpha$. For $\alpha = 0$, the composite stimulus has the identity of s_1, and for $\alpha = 1$ it has the identity of s_2. For intermediate values of α, a sharp transition between the two phonetic categories is observed. Identification and discrimination tests show that this relative intensity continuum is categorically perceived, and the discrimination model derived earlier is shown to adequately account for the observed discrimiantion results. Results of a selective adaptation test are also presented in which the composition of the adaptor is systematically varied along this relative intensity continuum. The results show that the boundary shifts are a strong function of the acoustic makeup of the adaptor. Dichotic listening results are also presented in which the effect of interaural intensity and stimulus composition are simultaneously investigated. The results indicate that this experimental paradigm should be useful for studying ear dominance. Lastly, a tentative model based on power law excitation of neural populations is investigated in an attempt to unify the results of the identification, discrimination, selective adaptation and dichotic listening tests.

ACKNOWLEDGEMENTS

I would like to acknowledge the help and support I received from the department of Linguists during the course of this thesis. Above all, I would like to gratefully acknowledge the generous and patient assistance of my subjects who, without the usual stimulant of financial reward, provided that data which forms the substantive part of this thesis. The quality of the results presented is entirely due to their conscientious performance in the laboratory.

I would like to thank the members of the faculty of the Department of Linguistics for their many helpful conversations and comments during the course of this research, and especially to Drs. J.T. Hogan, T.M. Nearey, and my supervisor, A.J. Rozsypal. The helpful comments and criticisms of Drs. R. Rink and J.H. Gilbert during the final examination are also gratefully acknowledged.

Lastly, I would like to thank the Department of Linguistics and the Canada Council for their support in the form of Doctoral Fellowship (Award No. 452-772845) for the period May 1977 to May 1978.

TABLE OF CONTENTS

CHAPTER PAGE

CHAPTER 1. INTRODUCTION

1.1 Background to Categorical Perception 1
1.2 Categorical Perception and Selective Adaptation 6
1.3 The Present Research 8

CHAPTER 2. MODELS OF CATEGORICAL PERCEPTION

2.1 Phonetic-memory Models 11
  2.1.1 The Effect of Step Size on Discrimination 14
  2.1.2 The AX Discrimination Test 20
  2.1.3 Phonetic vs. Auditory Discrimination 27
2.2 Signal Detection Models of Categorical Perception 29
  2.2.1 Dispersion 31
  2.2.2 Dispersion and AX Discrimination 37
  2.2.3 Finding a Dispersion Function 45
2.3 A Detector Model of Categorical Perception 46
  2.3.1 The Dispersion Function 53
  2.3.2 Calculating the AX Discrimination Function 59
2.4 Application to Selective Adaptation 61
  2.4.1 Discrimination under Conditions of Adaptation 63
  2.4.2 The Effect of Adaptation on the Stimulus Trajectory 64
  2.4.3 A New Experimental Paradigm 67
2.5 Summary 69

CHAPTER 3. CATEGORICAL PERCEPTION OF A RELATIVE INTENSITY CONTINUUM

3.1 Creation of the Test Stimuli 74
3.2 Identification curves 82
  3.2.1 Experimental Setup 82
  3.2.2 Experiment 1:Identification Scores 90
  3.2.3 Experiment 2:Identification of /bɛt/-/dɛt/ 99
  3.2.4 Experiment 3:Identification of Liquids and Vowels .. 101
  3.2.5 Experiment 4:Identification of /ba/-/da/-/ga/ 103
  3.2.6 Summary of Identification Tests 110
  3.2.7 Experiment 5:Masking by a Vocalic Masker 112
3.3 ABX and AX Discrimination Tests 116
  3.3.1 Experiment 6:ABX Discrimination 116
3.4 AX Discrimination 120
  3.4.1 Experiment 7:AX Discrimination Scores 122
  3.4.2 Fitting the AX Discrimination Data 129
  3.4.3 Summary of Discrimination Results 133
  3.4.4 What is Creating the Dispersion? 135

CHAPTER 4. SELECTIVE ADAPTATION

4.1 EXPERIMENT 8:Selective Adaptation 139
4.2 Summary 147
4.3 EXPERIMENT 9:Effect of Adaptor Intensity 148

CHAPTER 5. MODELLING THE EXPERIMENTAL RESULTS

5.1 A Preliminary Model of /b/-/d/ Detection 155
5.2 Intensity Coding by Neural Populations 161
5.3 Inhibition Between Neural Populations 166
5.4 Fitting the AX Discrimination Data 175
5.5 Modelling Selective Adaptation 180
  5.5.1 Comparison with Keeler's (1968) Model 186
  5.5.2 Modelling the Selective Adaptation Paradigm 188
  5.5.3 Fitting the /dae/-Adaptor Data 190
  5.5.4 Fitting the /bae/-/dae/ Adaptor Data 196
5.6 Summary 200
5.7 Extension to Dichotic Listening 202

CHAPTER 6. BINAURAL FUSION

6.1 EXPERIMENT 10:Dichotic Presentation 204
6.2 Ear Dominance 213
6.3 A Model of Binaural Interaction 220
  6.3.1 Calculating Ear Dominance Curves 223
  6.3.2 Predicting the Category Boundaries 233
6.4 Predicting the Dichotic Responses 236
6.5 Summary 239

CHAPTER 7. SUMMARY AND CONCLUSIONS

7.1 Summary 241
7.2 Directions for Further Research 250

REFERENCES.. 256

LIST OF FIGURES

FIGURE

2.1 Hypothetical Identification Curve 16
2.2 Calculated ABX Response Surface 17
2.3 ABX Discrimination Curves for Various Step Sizes 19
2.4 Calculated AX Response Surface 23
2.5 AX Discrimination Curves for Various Step Sizes 25
2.6 Hypothesized AX Discrimination Function 26
2.7 SDT Model of AX Discrimination 32
2.8 Effect of Dispersion on a Perceptual Continuum 35
2.9 SDT Model of AX Discrimination with Dispersion 38
2.10 AX Discrimination Curves for Various Step Sizes 40
2.11 Comparison of Phonetic-Memory and Dispersion Model 42
2.12 Gaussian Detector Response Functions 50
2.13 Two-Dimensional Decision Plane 50
2.14 Stimulus Trajectory in the Decision Plane 52
2.15 Various Dispersion Curves 56
2.16 AX Discrimination Surfaces for Gaussian Detectors 60
2.17 Effect of Adaptation on the Stimulus Trajectory 65
2.18 Hypothetical Stimulus Trajectory for Composite Stimuli 70

3.1 Schematic Formant Transitions 75
3.2 Covariance traces for alignment of stimuli 75
3.3 Aligned /b/ and /d/ Formant Transitions 78
3.4 Variation in Amplitude with α 80
3.5 Sonagrams of Composite /bae/-/dae/ stimuli 84
3.6 Computerized Presentation Facility 85
3.7 Linearity of Braun Amplifiers 87
3.8 Frequency Response of Headphones 88
3.9 /bae/-/dae/ Identification Curves 92
3.10 Long-term Stability of /bae/-/dae/ Boundary 95
3.11 Extreme Limits of /bae/-/dae/ Boundary 97
3.12 /bɛt/-/dɛt/ Identification Curves 100
3.13 /ra/-/la/ Identification Curves 100
3.14 Hypothetical /b/-/d/-/g/ Decision Space 105
3.15 Summary of /b/-/d/-/g/ Identification Results 107
3.16 /b/-/d/-/g/ Identification Curves 108
3.17 Graphical Interpretation of /b/-/d/-/g/ Curves 109
3.18 /bae/-/dae/ Identification in Presence of /ae/ Masker 114
3.19 ABX Discrimination Curves 118
3.20 AX Discrimination Curves 125
3.21 Summary of AX Discrimination Curves for /dae/ Standard 128
3.22 Best Fit of AX Discrimination Model 132
3.23 Dispersion Function for Best Fit Model 134

4.1 Hypothetical Boundary Shifts 140
4.2 Pre- and Post-Adaptation Identification Curves 142
4.3 Boundary Shift as a Function of Adaptor Composition 142
4.4 Variation of Width of Identification Boundary 146
4.5 Boundary Shifts as a Function of Adaptor Intensity 150
4.6 Comparison of Mixed- and Pure-Adaptor Boundary Shifts 150
4.7 Hypothesized Boundary Shifts for /bae/-/dae/ Adaptor 152

5.1 Stimulus Trajectory for θ=0.27 157
5.2 Dispersion Function for θ=0.27 157
5.3 Stimulus Trajectories for Various Values of θ 160
5.4 Schematic Model of Neural Excitation 163
5.5 Schematic Model of Inhibiting Populations 167
5.6 State Diagram for Inhibiting Populations 167
5.7 Isoclines for Inhibition Model 171
5.8 Dispersion Curves for n=1 Model 173
5.9 Dispersion Curves for n=2 Model 174
5.10 Calculated AX Discrimination Curves 178
5.11 State Diagram for Inhibition and Adaptation 182
5.12 Fit of Model to /dae/-Adaptor Data 191
5.13 Corrected Fit of Model to /dae/-Adaptor Data 194
5.14 Calculated Boundary Shifts for Composite Adaptor 197
5.15 Corrected Boundary Shifts for Composite Adaptor 199

6.1 Schematic Dichotic Presentation of /bae/-/dae/ Stimuli 206
6.2 Dichotic /bae/ Identification Curves 210
6.3 Typical /bae/ Identification Curves 211
6.4 Averaged /bae/ Identification Curves 212
6.5 Ear Dominance Curves for /bae/ and /dae/ Stimuli 215
6.6 Efron and Yund Model of Ear Dominance Curves 216
6.7 Calculated Ear Dominance Curves 225
6.8 Calculated Stimulus Trajectories 225
6.9 Models of Peripheral /b/-/d/ Detection 227
6.10 Ear Dominance Curves for Coupled Detectors 229
6.11 Stimulus Trajectory for Coupled Detectors 229
6.12 Best Fit Ear Dominance Curves 232
6.13 Calculated /bae/ Identification Curves 238

7.1 Schematic Model of /b/-/d/ Detection 246
7.2 Hysteresis and Identification Curves 254

LIST OF TABLES

TABLE

2-1 Response Probabilities for the 2IAX Paradigm 21

4-1 Pre- and Post-Adaptation Boundaries 143

6-1 Ear and Stimulus Dominance Indices 219
6-2 Ear Dominance Weighting Factors 231
6-3 Predicted /bae/-/dae/ Boundaries 234
6-4 Ear Dominance Factors 235

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND TO CATEGORICAL PERCEPTION

Categorical perception in speech was originally characterized by Liberman, Harris, Hoffman and Griffith (1957) in a classic experiment in which it was shown that the discriminability of a series of synthetic /b/-/d/-/g/ stimuli was poorer within phonetic categories than between categories. The outcome of this study is well known, and established a methodology for investigating categorical perception which has lasted two decades. Liberman et al. (1957) constructed a preliminary model of the ABX -

discrimination results based on the extreme assumption that the ability to discriminate between stimuli was strictly a result of overt phonetic classifications of stimuli. This model became the standard "test" for categorical perception: if the discrimination results could be adequately predicted by this model, then the continuum was "categorically perceived" (Studdert-Kennedy, Liberman, Harris and Cooper, 1970). In the intervening years since the Liberman et al. study, various continua, both speech and nonspeech, have been shown to be categorically perceived. However, this proliferation of categorized continua has not brought with it a deeper understanding of the nature of categorical perception itself.

Discrimination studies invariably showed that the data and the Liberman et al. model consistently differed: discrimination, as measured by the ABX paradigm, was better for within-category stimuli than was predicted on the basis of category labelling. Fujisaki and Kawashima (1969, 1970) extended the ABX discrimination model to admit an element of auditory discrimination, i.e., discrimination based on auditory and not phonetic characteristics of the signal. Their model added an additional parameter to the Liberman et al. model (which has come to be known as the "Haskin's model"), and consequently improved the fit between model and data. The assumption underlying their model was that auditory memory decays faster than phonetic memory, with phonetic categorization playing the major role. The Fujisaki and Kawashima model, while improving the fit between the model and the data, did not substantially increase the understanding of the phenomenon itself, since the model did not require as input any physical specification of the stimuli which were being discriminated.

The criteria for the demonstration of categorical perception were canonized by Studdert-Kennedy et al. (1970). In their formulation, the discrimination results were required to show an enhanced peak at the phoneme boundary (the "phoneme boundary effect"), and this peak had to be predicted on the basis of the labelling probabilities. Using these criteria, various researchers were subsequently able to show that various nonspeech continua were also categorically perceived (e.g., Locke and Kellar, 1873; Cutting and Rosner, 1974; Pastore, Ahroon, Baffuto, Friedman, Puleo and Fink, 1976). The belief now commonly held is that categorical perception may be a phenomenon characterizable at the psychophysical level in which the acoustic structure of the stimuli plays a direct role (Pastore et al, 1976; Miller, Wier, Pastore, Kelly and Dooling, 1976; Carney, Widin and Viemeister, 1977; Cutting and Rosner, 1974). Part of this change of view evidently stems from the demonstration of extant "categories" in neonates (Eimas, Siqueland, Juszyck and Vigorito, 1971; Morse, 1972; Kuhl and Miller 1975a) and animals (Kuhl and Miller, 1975b). The question which arises from these studies is whether or not categorical perception results from natural perceptual boundaries or learned category boundaries. Evidence exists to show that both may be involved. The demonstrations with chinchillas indicate that certain speech stimuli (e.g., /d/ and /t/) are sufficiently far apart in perceptual space as to be readily associated with events in the environment (Kuhl and Miller, 1975b), suggesting that /d/ and /t/ are "natural" categories. On the other hand, Miyawaki, Strange, Verbrugge, Liberman, Jenkins and Fujimura (1975), on the basis of /r/-/l/ distinctions of adult Japanese and Americans, show that the /r/-/l/ continuum is categorically perceived by English subjects but not so perceived by adult Japanese. The implication of this study is that some instances of categorical perception in speech may be attributable to learned distinctions. There is no evidence to date which clearly suggests that all instances of categorically perceived continua are attributable to the same underlying mechanisms.

Several mechanisms for categorical perception recently have been proposed which are plausible in their description but rather vague in their formulation. Miller et al. (1966) suggest that it is

> "... a _single component_ or a _stimulus complex_ that

> is the variable. It is likely that the unchanged or constant part of the stimulus complex provides an immediate stimulus context against which the effects of the changed component are judged." (p. 415)

while Pastore et al. (1976) suggest that categorical perception is due to

> " ... a single, sharp, stable dichotomy or limitation along a dimension causes both a natural tendency to form a category boundary and, at the same time, improves the precision of the information used in discriminating stimuli separated by the dichotomy or limitation." (p.694)

However, neither Miller et al. nor Pastore et al. formalize their models, which makes them difficult to test. Their intent is obvious, however: categorical perception may reflect processing limitations of the sensory systems. This appears especially true for the Pastore et al. (1976) experiment in which critical flicker fusion (CFF) was shown to be categorically perceived along the frequency-of-flicker dimension.

Categorical perception, then, remains an enigma. It is fairly easily demonstrated, perhaps too easily, but has yet to be explained in any psychophysical sense. While various proposals, such as rapid decay of auditory information with slow decay of phonetic information (Fujisaki and Kawashima, 1969; Pisoni, 1975), fixed signal components (Miller et al., 1976) or a "stable dichotomy" (Pastore et al., 1976), are reasonable, they have not been formalized into models which require as input some physical variable associated with the stimuli.

1.2 CATEGORICAL PERCEPTION AND SELECTIVE ADAPTATION

Since Eimas and Corbit (1973) it has been often suggested that extraction of phonetic information from speech is mediated by neural constructs called "feature detectors". These feature detectors presumably span a stimulus continuum[1], and are characterized by a response function which represents the sensitivity of the detector to stimuli along this continuum. Under repeated presentation of a stimulus drawn from this continuum, the detector is thought to become desensitized, with the result that the stimulus value for which the two detector outputs are equal shifts in the direction of the adapting stimulus. This shift has been taken as support for the notion that the continuum is spanned by two separate neural entities which selectively respond to stimuli along the continuum. As in the case of categorical perception, interpretation of phonetic boundary shifts is complicated for want of a quantitative model. With the exception of Elman (1979), no explicit formulation of a detector model has yet been constructed. Elman's model assumes two Gaussian detector response functions, and with this model he shows that it is possible to account for boundary shifts by a change in the subject's response bias. However, the model has not been investigated in sufficient detail to show that this is the only way in which the model can account for these shifts.

---

[1] A "continuum" is a physical parameter of the stimuli which, when varied over a certain range, causes a change in percept from one phonetic category to another.

Now, what is the relation of selective adaptation to categorical perception? These two phenomena have traditionally formed two separate lines of research in the speech perception literature, but it seems clear that they must be related. Cooper (1974) investigated the change in ABX discrimination under adaptaion, and his results indicate that the peaks of the discrimination curves shift in accordance with the shifts in the category boundaries of the labelling curves. This suggests that (a) identification and discrimination involve the same physiological/perceptual mechanisms, and (b) selective adaptation affects this system in such a way that the peak of discriminability tends to follow the category boundary. (It is not certain from his results that the discrimination peak and the labelling boundary always coincide, but this is a distinct possibility). It remains to be shown, however, that a two-detector configuration spanning some physical continuum can simultaneously account for categorical perception of the continuum as well as the boundary shifts under selective adaptation.

The detector theory of speech perception received considerable impetus from selective adaptation studies, but the concept of "detector" itself has a rather uncertain status. Usually it reduces to little more than a graphical construct to aid in the interpretation of results (e.g., Miller 1977; Ainsworth, 1977). The failure to give mathematical form to the theory weakens it rather than stengthens it, since arguments for and against this interpretation of category boundary shifts reduce to exercises in verbal logic. The reluctance to formalize the detector model perhaps stems from the fear of increasing the vulnerability of the model by making it more explicit. Even though detectors are at present little more than "physiological metaphors" (Simon and Studdert-Kennedy, 1978), a translation into a mathematical metaphor is certainly desirable. The quality of the metaphor is then relatable to how well it can quantify the phenomena it is supposed to explain.

1.3 THE PRESENT RESEARCH

This sets the stage for the present research. First, various models of discrimination (and hence categorical perception) are discussed in Chapter 2, and lead to a formulation of a signal detection theory (SDT) model of discrimination. This model is based strictly on auditory discrimination, and centers around the concept of "dispersion", which reflects a non-linear mapping between the physical acoustic continuum and the corresponding perceptual continuum. It is shown that a detector model along the lines of Eimas and Corbitt (1973) and as formulated by Elman (1979) is a dispersive system, and theoretically can account for the categorical perception of a continuum. It is also shown that such a detector configuration simultaneously can account for phonetic boundary shifts under adaptation. (This is the same model which Elman uses to support a response bias account of selective adaptation).

In Chapter 3 a new experimental paradigm is described for investigating categorical perception and various other speech phenomena. Briefly, two CV syllables (e.g., /bae/ and /dae/) are mixed together by adding their digitized waveforms. A categorically perceived continuum is then formed when the relative intensities of the two component signals is varied. Various experiments are described which investigate the generality of this form of categorical perception. In Chapter 4, a selective adaptation experiment is described in which boundary shifts are shown to be a strong function of the relative intensities of the /bae/ and /dae/ components of the adaptor. These results suggest that /b/ and /d/ detectors are effected independently and simultaneously by the two components of the signal.

In Chapter 5, a model of this "monaural fusion" paradigm is constructed which assumes that the /b/ and /d/ processors are functionally separate, i.e., the /b/ component of the stimulus is recognized by a processor which recognizes /b/, and the /d/ component is recognized by a processor which recognizes /d/. Following the basic framework laid out in Chapter 2, the model is extended to account for discrimination and selective adaptation, and suggests that for all intents and purposes, the two signal components behave as if they were presented over separate auditory channels. A binaural extension of the model in Chapter 6 is used to interpret the results of a dichotic listening experiment, and the application of the model to these results suggests tha there is only a slight coupling of the /b/ and /d/ processors. The four experimental paradigms - identificaton, discrimination, selective adaptation and binaural fusion are thus shown to be interpretable by a single model, a direct consequence of the fact that this relative intensity continuum is categorically perceived.

CHAPTER 2

MODELS OF CATEGORICAL PERCEPTION

2.1 PHONETIC-MEMORY MODELS

Categorical perception is defined by two observable measures: the identification function (representing a subject's ability to label stimuli which differ along some physical continuum), and the discrimination function (representing a subject's ability to discriminate between stimuli drawn from that continuum). The criteria for demonstration of categorical perception are stated in terms of these two psychometric functions (Studdert-Kennedy et al., 1970), and the "test" for categorical perception is how well discrimination can be predicted from the corresponding identification curve. The Haskins model (see Chapter 1) has been shown to generally underpredict the discriminability of most speech continua, and has been modified to include a measure of auditory discriminability by Fujisaki and Kawashima (1969, 1970). Both of these models assume that discrimination of speech stimuli presumably results from an overt phonetic classification by the subject, and will be referred to hereafter as "phonetic memory models".

In the Liberman et al. (1957) model of ABX discrimination, the results of an identification, or labelling, test are used as a posteriori estimates of the probability that a subject will perceive a stimulus as belonging to one of the phonetic categories. The expected discrimination scores are then computed by enumerating the various response probabilities for the ABX test paradigm using these labelling probabilities (see MacMillan et al., 1977; Pollack and Pisoni, 1971). For instance, if p_1 and p_2 are the probabilities that two stimuli x_1 and x_2 contrasted in the ABX[1] paradigm are classified by the subject as, say, category B, then the predicted discrimination score is given by

---

[1] In the ABX paradigm under discussion, X is always either A or B.

$$P_D = 0.5[1 + (p_1 - p_2)^2] \qquad (2\text{-}1)$$

Assuming the identification function shows perfect within category labelling, the probability of discriminating two within-category stimuli near the endpoints of the continuum (i.e. $p_1 = p_2 = 0$ or $p_1 = p_2 = 1$) is

$$P = 0.5 \qquad (2\text{-}2)$$

That is, discrimination should occur at a strictly chance level. Experimental results show, however, that some within-category discrimination is possible. To accommodate this disparity, Fujisaki and Kawashima (1970) extended the Haskin's model by positing a two-tier discrimination process. Discrimination between stimuli, as in the Haskin's model, is considered to be an operation primarily involving explicit phonetic categorization. If a subject perceives the stimuli in contrasting phonetic categories, he responds accordingly. However, if the subject does not perceive the stimuli in contrasting categories, he then attempts to discriminate by comparing the "timbres" of the auditory images. The measure of auditory discriminability is represented in the model by a "guessing factor" T. The resulting equation is (Fujisaki and Kawashima, 1970)

$$P_D = 0.5[(p_1 - p_2)^2 + p_1(1-p_2) + p_2(1-p_1)] + T[p_1 p_2 + (1-p_1)(1-p_2)] \qquad (2\text{-}3)$$

When T=0.5, this Fujisaki and Kawashima model[2] reduces to the Haskins model. These models have been used extensively to predict discriminability in studies of categorical perception. Since ABX discrimination experiments invariably show greater than chance discriminability for within-category stimuli, the FK model shows a superior fit since T corresponds essentially to a "d-c shift" of the predicted discrimination curve.

2.1.1 The Effect of Step-size on Discrimination

It is traditional in ABX discrimination studies to perform the experiment using "one-step" and/or "two-step" intervals. The companion identification test is carried out using a set of stimuli x_i, i = 1,2,3, ..., n, where all stimuli are separated by Δx along the physical stimulus continuum. The ABX test is carried out using stimulus pairs separated by either Δx or $2\Delta x$. The two-step test invariably shows an overall increase in discriminability over the one-step test, as well as broader peaks.

---

[2] Hereafter referred to as the "FK" model.

The influence of step size on discriminability has not recieved a great deal of attention in the literature. The notions of "one-step" or "two-step" are not sufficiently well-defined to constitute any sort of standard since the step size itself is arbitrary. To see the influence of step size, Equation 2-3 above can be calculated for all combinations of two stimuli x_1 and x_2 separated by a constant amount Δx. For purposes of illustration, the identification is assumed to be given by the normal ogive

$$p(x) = \Phi\left(\frac{x-0.5}{\sigma}\right) \qquad (2\text{-}4)$$

where x is a hypothetical stimulus continuum ranging from 0 to 1 (see Fig. 2.1) and p(x) is the probability that a stimulus x will be classified as, say, category B. Φ is the normal cumulative distribution function. The width of the transition region (as characterized by σ) is arbitrarily set at $\sigma=0.05$. Using the identification function given by Equation 2-4 and a step size of 0.05, $P_D(x_1,x_2)$ can be computed for all possible pairings of the stimuli.[3] The result is a three-dimensional response surface as shown in Fig. 2.2. The two dimensions in the horizontal

---

[3] For the rest of the discussion, the quantity $P_D(x_1,x_2)$ will be referred to as the "discrimination function" and p(x) will be referred to as the "identification function" or "labelling function".

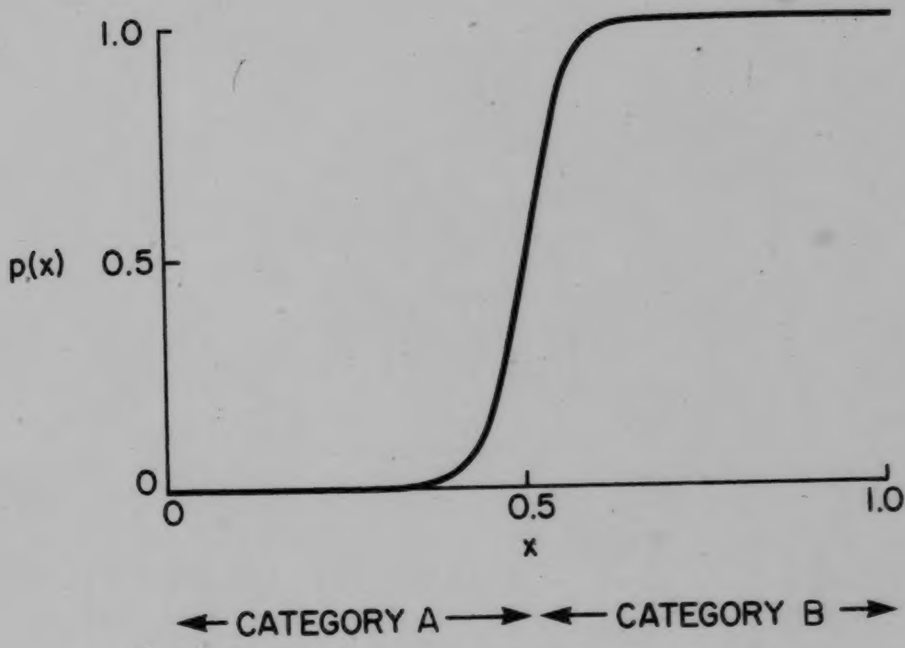


Fig. 2.1. Hypothetical identification function (normal ogive with mean 0.5 and standard deviation 0.05) used in the calculation of the discrimination functions described in the text

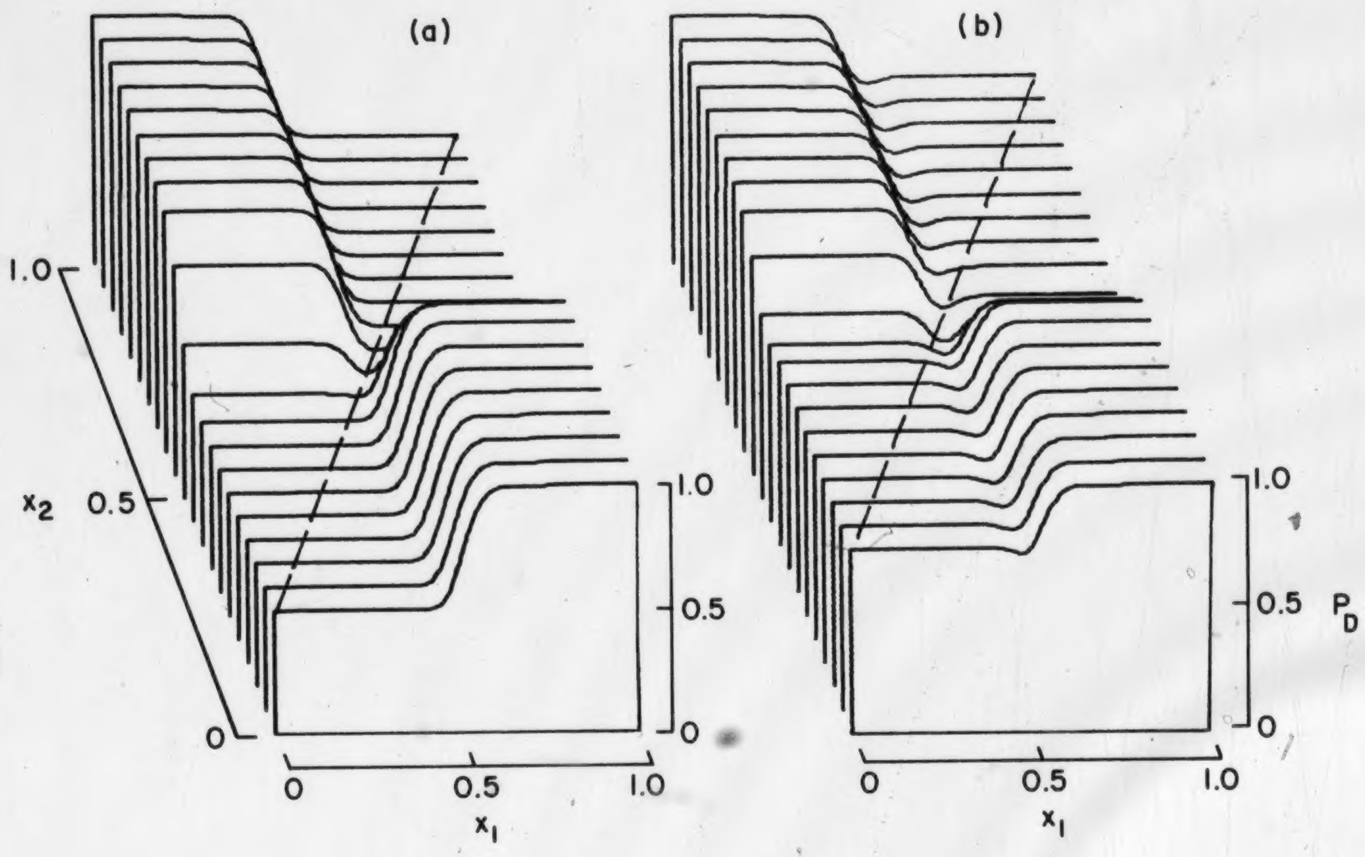


Fig. 2.2. Calculated response surfaces $P_D(x_1,x_2)$ for the ABX paradigm. (a) is the Haskins model (T=0) and (b) is the Fujisaki and Kawashima model with T=0.2. The dashed lines represent the zero step size condition, $x_1 = x_2$

plane represent the positions of x_1 and x_2 along the x-continuum, and the vertical axis (P) represents the probability of discriminating x_1 and x_2. Profiles of constant x_2 represent a hypothetical fixed-standard ABX test, while sections through this surface parallel to the diagonal dashed line represent a variable-standard fixed step size ABX test (i.e., the usual ABX paradigm). The FK model (Equation 2-3) predicts that in case the two stimuli being contrasted are physically identical, better than chance discriminability is predicted (for $T > 0$). In the vicinity of the boundary it decreases slightly (see Fig. 2.3 b). Another novel feature is the slight dip on each side of the peak of the discrimination curve (Fig. 2.3b).

This model of the ABX discrimination process, while capturing the general shape of observed discrimination curves, suffers from a rather undesirable limiting behaviour. It seems counter-intuitive that physically identical stimuli can be discriminated at better than chance level. Furthermore, that the discrimination curve should drop at the phonetic boundary for small step sizes is also disconcerting. Only the Haskins model (Fig. 2.3a) demonstrates the correct limiting behaviour since for zero step size, strictly chance discriminability is predicted at all points along the continuum. These properties of the FK model have remained obscured since virtually all applications of the model to date have involved the use of

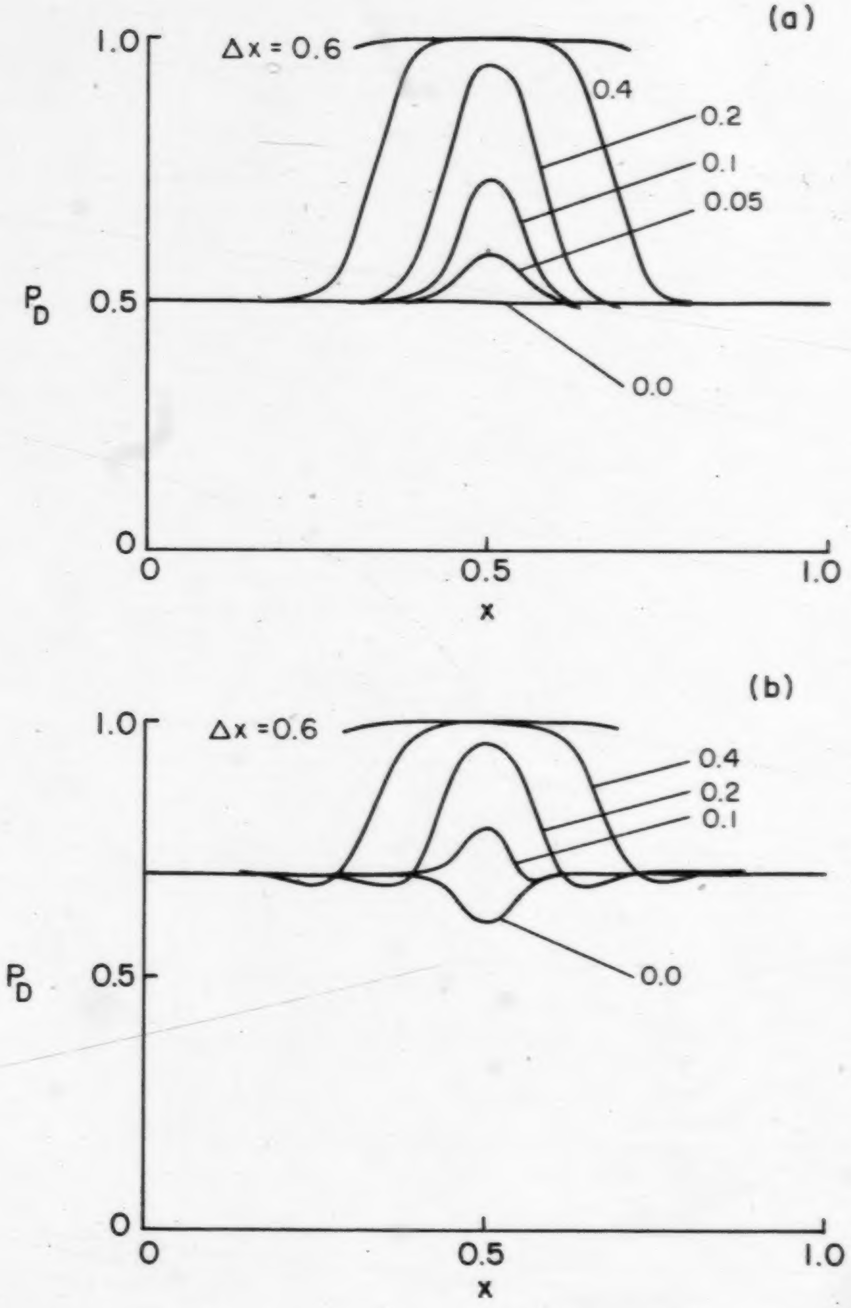


Fig. 2.3. Calculated ABX discrimination functions for various step sizes, Δx. (a) Haskins model (b) Fujisaki and Kawashima model with T=0.2

experimentally determined relative frequencies. Since identification functions are sampled only at a few intervals along the continuum (typically only 10 or so) and often show considerable statistical scatter, predicted discrimination curves do not show these minor effects.

2.1.2 The AX Discrimination Test

The ABX paradigm unquestionably has been the most popular discrimination test in studies on categorical perception. Until recently, the AX paradigm has received less attention. The reasons for its lack of popularity are unclear, but Zinnes and Kurtz (1968) attribute it to

> "... the very old belief that the 'same' or 'equal' category in discrimination experiments is too unstable, too easily influenced by the subject, and as such gives a poor measure of a subject's optimum discriminability." (p. 392).

However, recent studies on categorical perception have tended to favour the AX discrimination test (e.g., Repp et al., 1978; Carney et al., 1977; Williams, 1977; Wood, 1976; Cutting, Rosner and Foard, 1976; Pisoni, 1973).

Following the strategy for the FK model, a phonetic-memory model of AX discriminability can be computed (cf. Pollack and Pisoni, 1971; Zinnes and Kurtz, 1968). Assume that the probability that two stimuli x_1 and x_2 will be categorized as belonging to, say, category B are $p_1 = p(x_1)$ and $p_2 = p(x_2)$. The subject may then perceive these two stimuli as AA, AB, BA or BB. If he perceives either AB or BA, he will presumably respond "different", and if he perceives either AA or BB he will be forced to discriminate on the basis of non-phonetic differences and, as in the case of the FK model, will respond "different" with probability T. The factor T therefore incorporates the true within category discriminability as well as the subject's response bias. Assuming equal a priori probabilities of presenting x_1 and x_2 in either order ($x_1 x_2$ or $x_1 x_2$), the various response probabilities are as shown in Table 2-1 below.

TABLE 2-1

RESPONSE PROBABILITIES FOR THE 2IAX PARADIGM

PERCEIVED CATEGORY	STIMULUS	
	$x_1 x_2$	$x_2 x_1$
A A	$p_1 p_2$	$p_2 p_1$
A B	$p_1 (1-p_2)$	$p_2 (1-p_1)$
B A	$p_2 (1-p_1)$	$p_1 (1-p_2)$
B B	$(1-p_1)(1-p_2)$	$(1-p_2)(1-p_1)$

Since stimulus combinations $x_1 x_2$ and $x_2 x_1$ occur with equal frequency, the resulting proportion of "different" responses is

$$P_D = p_1(1-p_2) + p_2(1-p_1) + T[p_1 p_2 + (1-p_1)(1-p_2)] \tag{2-5}$$

This equation is similar in form to Equation 2-3 and has similar properties. In this case, T is more readily interpreted as a criterion which can be manipulated by the subject, as well as a factor which reflects enhanced discriminability for within-category comparisons. If the subject chooses to ignore subtle differences between stimuli or, alternatively, is unable to perceive any differences, then T=0 and Equation 2-5 becomes

$$P_D = p_1(1-p_2) + p_2(1-p_1) \tag{2-6}$$

which is the result derived by Zinnes and Kurtz (1968, p. 397).

This model of the AX discrimination test also shows incorrect limiting behaviour. When the two stimuli are physically identical, $p_1 = p_2$ and Equation 2-5 becomes

$$P_D = T + 2(1-T)(p_1 - p_1{}^2) \tag{2-7}$$

which predicts non-zero discriminability between physically identical stimuli. This can be seen in Fig. 2.4 where Equation 2-5 is calculated for all combinations of x_1 and x_2. Sections through the response surface for constant step

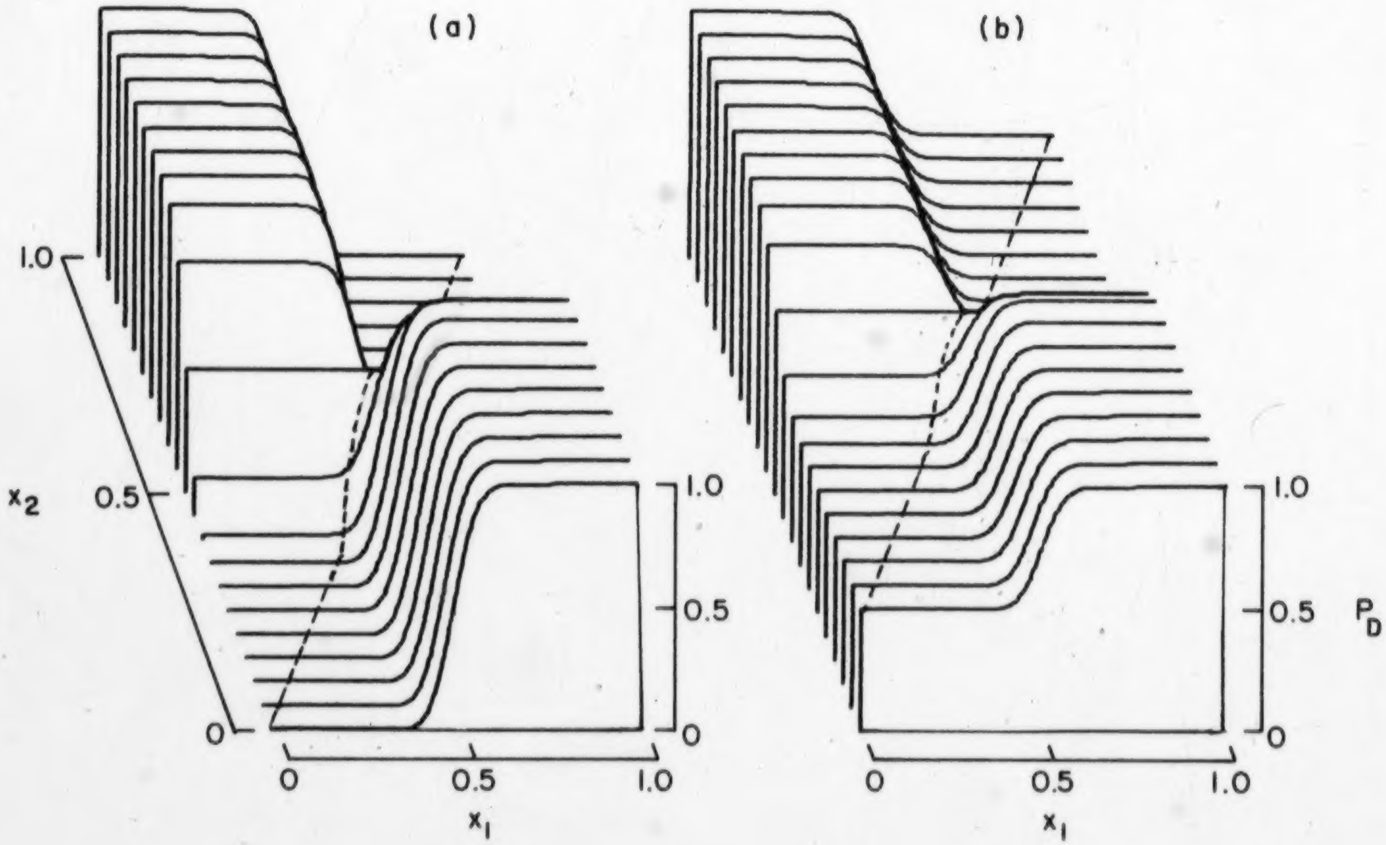


Fig. 2.4. Calculated discrimination response surfaces for the AX paradigm. (a) represents the case of T=0 (i.e., purely phonetic discrimination). (b) represents T=0.5. The dashed lines represent the zero step size condition, $x_2 = x_1$

size (i.e., sections corresponding to $x_2 = x_1 + \Delta x$) are shown in Fig. 2.5. This model also shows the curious prediction that when the standard stimlus is the boundary stimulus ($x = 0.5$ in this case), the discriminability is constant across the entire continuum. That is, when $p(x\ ) = 0.5$ Equation 2-7 reduces to

$$P_D = 0.5(1 + T) \qquad (2\text{-}8)$$

which is independent of the test stimulus or, for that matter, the entire test continuum.

Few experimental data are available to confirm or disconfirm predictions of this model. Variable standard (fixed step size) AX tests generally show a prominent peak in the vicinity of the phonetic boundary (cf. Foreit, 1977; Hanson, 1977), as this model predicts. With the exception of Carney et al. (1977), fixed-standard AX discrimination data have not been presented in detail in the literature. The Carney et al. data, however, show a considerable departure from the results predicted by the AX phonetic memory model (Equation 2-5). Carney et al. hypothesized that for a standard stimulus intermediate to the end-point and the boundary the discrimination curve ought to appear as shown in Fig. 2.6, and their results tend to support this view. These data, although limited, indicate a major failing of the phonetic-memory AX discrimination model.

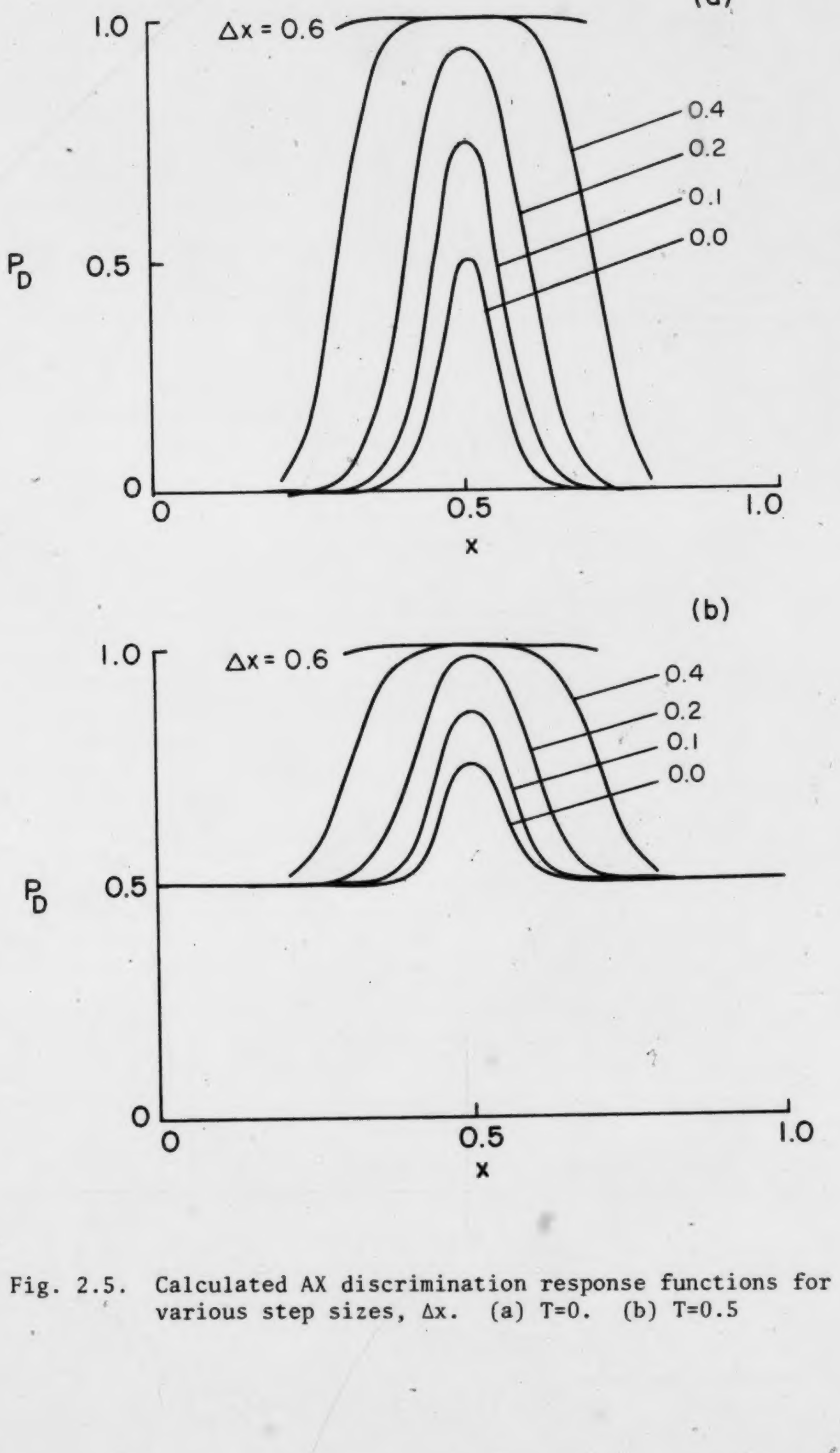


Fig. 2.5. Calculated AX discrimination response functions for various step sizes, Δx. (a) T=0. (b) T=0.5

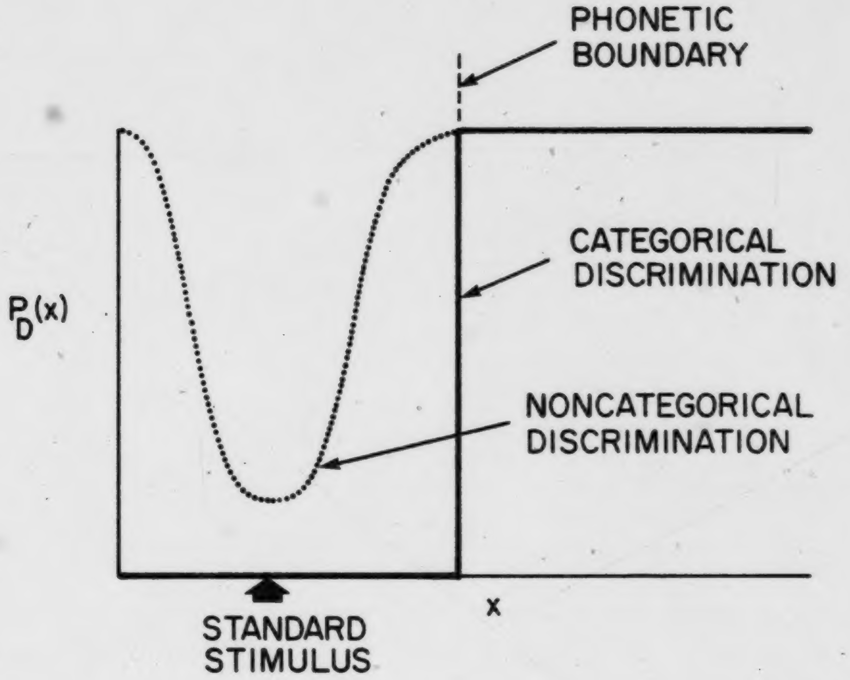


Fig. 2.6. Hypothesized AX discrimination function for the fixed-standard paradigm (after Carney et al., 1977). The heavy solid line represents what would be perfectly categorical discrimination.

Taken together with the improper limiting behaviour for zero step size, it seems quite clear that this model cannot be entirely correct in its formulation.

2.1.3 Phonetic vs. Auditory Discrimination

Whereas it was originally held that categorical perception was entirely a property of speech (see Pastore, 1976, for a historical overview), in recent years the tendency more and more has been to view categorical perception as resulting from psychophysical processes which may have little or nothing to do with the fact that the stimuli are speech-like. Carney et al. (1977), summarizing the results of their VOT study, suggest that

> "... auditory rather than exclusively phonetic explanations are the more appropriate." (p. 969).

Wood (1976) concludes that since his subjects did not overtly recognize short /pa/ and /ba/ stimuli as linguistic sounds and yet could successfully discriminate between them, the role of phonetic categorization in the determination of the "phoneme boundary effect" has perhaps been overplayed. Categorical perception experiments involving non-speech continua (e.g., Miller et al., 1976; Pastore et al., 1976; Cutting and Rosner, 1974; Locke and Kellar, 1973) as well as "phonetic categorization" by chinchillas (Kuhl and Miller,

1975) also suggest that categorical perception may reflect a more fundamental psychophysical phenomenon.

Few attempts at mathematical modelling of categorical perception have been carried out which assume auditory discriminability rather than phonetic discriminability, although various suggestions for models have been made (Miller et al., 1976; Pastore et al., 1976; Anderson et al., 1977). Anderson et al. (1977) present a neural model which has behaviour reminiscent of the observed results in discrimination studies. The complexity of their model, however, makes it difficult to apply to speech data. Miller et al. (1976) espouse the notion that there is something in the signal itself which results in categorical perception. Specifically, a stimulus from a categorically perceived continuum is supposed to contain a signal component against which the rest of the signal is compared. Categorical perception is viewed as a result of a masked threshold created by this contrast of signal components. Pastore et al. (1976) adopt a somewhat similar point of view, and propose that categorical perception arises from a

> "... common factor [which] involves either an internal (e.g., sensory threshold) or an external (e.g., a reference or interfering stimulus) limitation, which is both stable and more precisely defined than the typical differential sensory aspect (i.e., difference limen) of the continuum under investigation." (p. 687)

However, they do not formalize their proposal, and it is not immediately apparent how this model is to be implemented.

2.2 SIGNAL DETECTION MODELS OF CATEGORICAL PERCEPTION

The application of signal detection theory (SDT) to the specific problem of categorical perception has been almost completely neglected, the work by MacMillan et al. (1977) being a notable exception. As yet, no AX or ABX model has been presented in the speech perception literature as a replacement for the Haskins / FK model of discrimination. Consequently, in this section, a model of the AX discrimination process[4] will be developed which utilizes the familiar concepts of signal detection theory (Green and Swets, 1966).

The phonetic-memory models described above assume an all-or-none kind of perception similar to low threshold theory (MacMillan et al., 1977). This corresponds to assuming that the perception of a signal results in an internal discrete random variable Y assuming a value of either 0 or 1, where 0 corresponds to one signal category and 1 to the other. The alternative possibility is to let Y be a continuous random variable. In this case, the physical continuum x is mapped onto a perceptual continuum y by the function

---

[4] A model for the ABX test will not be attempted since more than one possible subject strategy is possible (see MacMillan et al., 1977; Pollack and Pisoni, 1971; Pierce and Gilbert, 1958).

$$Y = g(x) + \varepsilon(0,\sigma^2) \qquad (2\text{-}9)$$

where $y = E(Y) = g(x)$ and ε is a normally distributed noise component of zero mean and variance σ^2. For the present purposes, the mapping $y=g(x)$ will be assumed to be linear, i.e., $g(x)=x$.[5] If the perceptual continuum y is divided into two regions by a criterion y_b, then the probability of an arbitrary stimulus being identified as, say, category B (i.e., $Y \geqslant y_b$) is

$$p(Y>y_b) = \int_{y_b}^{\infty} \phi(y,\xi)\, d\xi \qquad (2\text{-}10)$$

The discrimination task can be modelled by computing the probability that Y_1 and Y_2 will be separated by some criterion Δy_c. The derivation of this model is straight forward (see Zinnes and Kurtz, 1968; Zinnes and Wolfe, 1977; Sorkin, 1962):

---

[5] This assumption, even if trivial, is necessary. Parameter x is a physical control variable which, if changed, causes a physical change in the stimulus. This change in the stimulus may or may not cause a corresponding change in the sensation variable y, depending on the function $y=g(x)$. In the analysis which follows, the independent variable will be shown as y, and it is implied that $y=x$.

$$P_D = 1 - \Phi\left(\frac{y_2 - y_1 + \Delta y_c}{2\sigma}\right) + \Phi\left(\frac{y_2 - y_1 - \Delta y_c}{2\sigma}\right) \qquad (2\text{-}11)$$

where Φ is the cumulative normal distribution. (Pierce and Gilbert, 1958, derived a similar model for AX discrimination, but only computed $P_D((y_1 - y_2) > \Delta y_c)$ rather than $P(|y_1 - y_2| > \Delta y_c)$. For purposes of comparison with the previously derived phonetic-memory model, y is assumed to range from 0 to 1 and the probability of discriminating two stimuli y_1 and y_2 can be computed from Equation 2-11.[6] The result of so doing is shown in Fig. 2.7 for two values of Δy_c. This model clearly cannot account for AX discrimination along a categorized continuum since it fails to demonstrate any enhanced discriminability in the vicinity of the boundary ($x = 0.5$). Along any line $x_1 = x_2 + \Delta x$, it can be seen that the predicted proportion of "different" responses is constant. As the criterion Δy_c is changed, the entire level of "different" responses increases or decreases by the same amount everywhere along this line. This model reflects "continuous" rather than "categorical" perception.

2.2.1 Dispersion

The principal failing of this model is the assumption that $y=g(x)$ is a linear mapping of the physical stimulus

---

[6] Since it was assumed above that $y = x$, the integration can be performed using y ranging from 0 to 1 instead of x from 0 to 1.

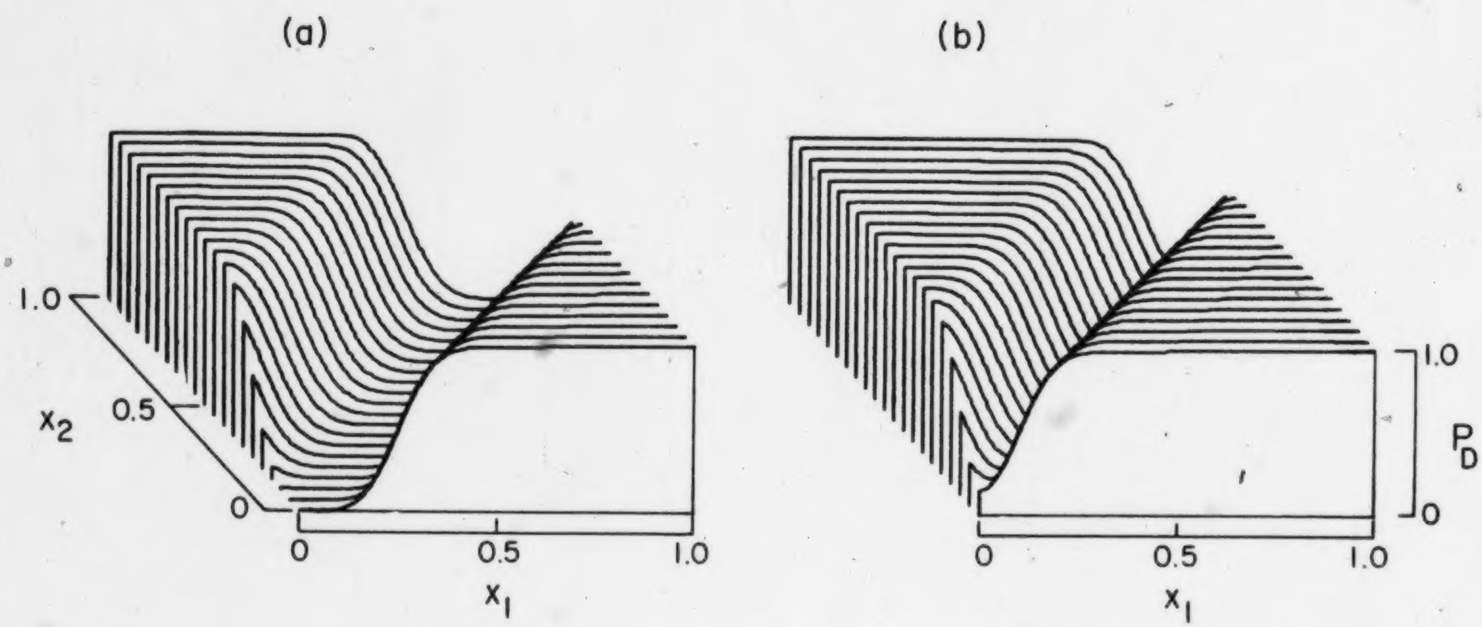


Fig. 2.7. SDT model of AX discrimination for the case of constant dispersion. (a) $\Delta y_c = 1.0$. (b) $\Delta y_c = 0.25$.

dimension x onto the perceptual dimension y. This assumed linearity, y=x, results in

$$D(x) = dy/dx = 1 \tag{2-12}$$

Stated otherwise, the dispersion dy/dx is constant.[7] Now, the existence of a peak in the discrimination curve in categorical perception studies indicates that the dispersion is not constant. Within-category stimuli are mapped onto the perceptual dimension y such that the distance between them is small, whereas stimuli near the boundary are separated by larger perceptual distances. In other words, the dispersion is greater in the vicinity of the phonetic boundary, and this enhanced dispersion effectively defines the boundary. For an arbitrary dispersion function D(x), the position of a stimulus x along the perceptual dimension y is

---

[7] "Dispersion" will be defined as the rate at which the perceptual variable y changes with respect to the physical variable x. As a physical analogy, consider a ray of light of wavelength λ passing through a refractive medium. The optical dispersion of the medium is given by $dN/d\lambda$, where N is the refractive index. If y is the position of the beam after passing through the medium, then because of the dispersion, y will change according to how fast N changes with λ. A medium for which $dN/d\lambda$ is non-zero is called a "dispersive medium".

$$y = \int_{x_0}^{x} D(\xi)\, d\xi \qquad (2\text{-}13)$$

where ξ is a dummy variable of integration (representing distance along the physical dimension x) and x_0 is some convenient reference point. Cast in these terms, the enhancement of discriminability at a phoneme boundary (or other perceptual boundary) can be represented by a peak in an underlying dispersion function. A peak in the dispersion function at x_0 will map all values of $x < x_0$ onto one end of the y scale, and all values of $x > x_0$ onto the other end of the y scale. This results in a spreading of the y-dimension with respect to the x-continuum in the vicinity of the catagory boundary, as is illustrated in Fig. 2.8b for the Gaussian dispersion function shown in Fig. 2.8a. (This choice of function is without theoretical import and is used for illustrative purposes only). Fig. 2.8c shows the same transformation in another form. It can be seen that there is a progression from one "state" to the other, with the steepness of the transition being inversely proportional to the width of the underlying dispersion function.

As the width of the dispersion function decreases to zero, the dispersion function approaches a delta function and y(x) becomes a unit step function. This represents perfect categorization, since y can only take on values of 0 or 1. Thus, purely phonetic categorization corresponds to infinite dispersion at the boundary. This is one extreme

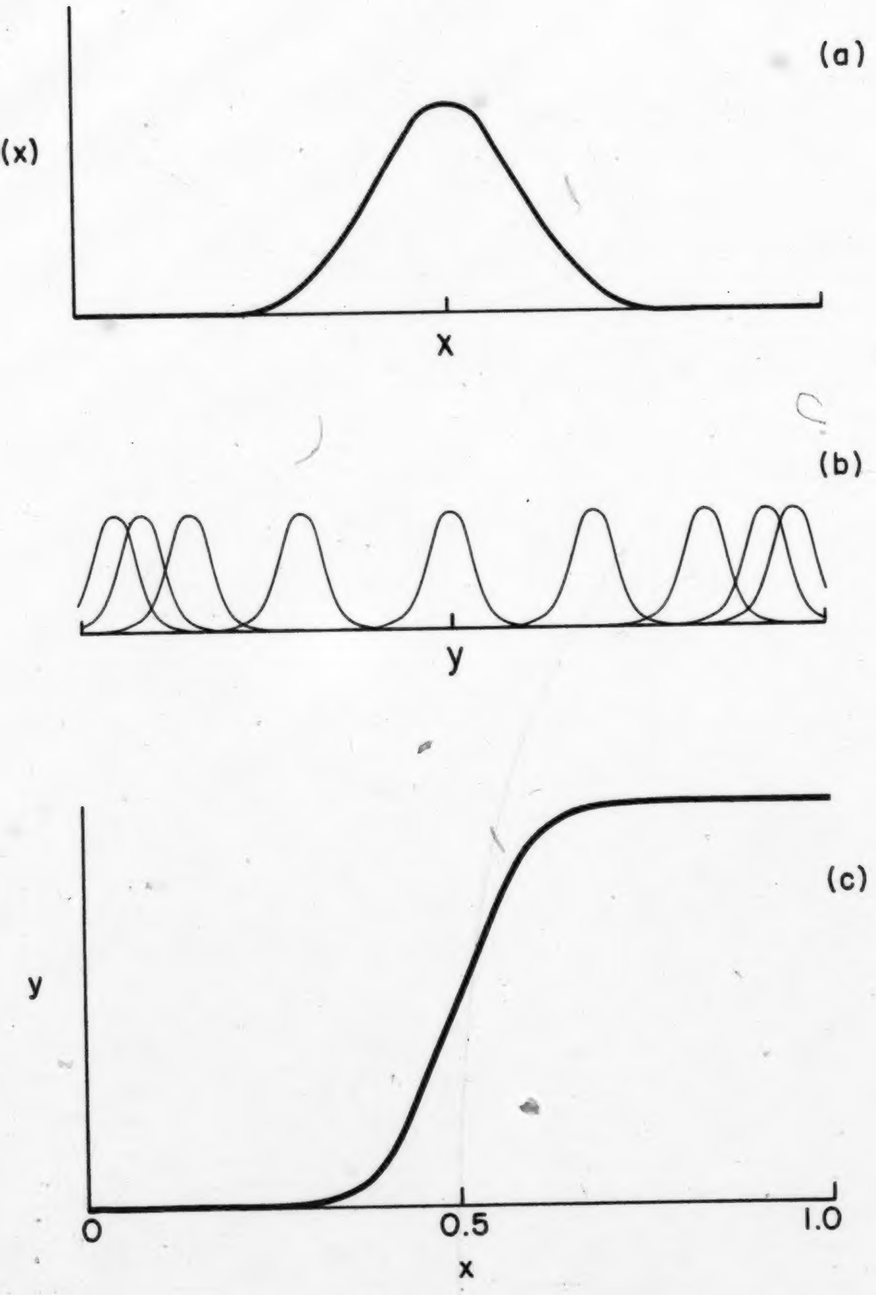


Fig. 2.8. (a) Gaussian dispersion function. (b) Effect of dispersion on the sensation continuum y. (c) The x → y mapping caused by the dispersion (a)

property of the model. Another extreme occurs when the dispersion is constant (already discussed above), and this corresponds to continuous perception. But, and this is the important point, depending on the width and shape of the dispersion function, various degrees of "categorical perception" are possible. It is therefore probably meaningless to insist on a dichotomous distinction between "categorical" perception and "continuous" perception, since perfect examples of either have yet to be found. It is probably more reasonable to view some continua as either "more categorical" or "less categorical" than others. Dispersion is a physical property of any signal processing system, although in most non-biological systems it is designed to be constant (i.e., some physical parameter maps linearly onto some measurable quantity, e.g., voltage).[8] The argument should not be whether or not dispersion exists, but rather how much dispersion exists.

Suggesting dispersion as a property of the receptive medium captures conceptually and mathematically what has been observed in discrimination studies all along: the organism does not respond equally to stimuli taken from different points on a continuum. Expressed differently, the sensation variable Y associated with a stimulus x in general need not change linearly with x. The nonlinear mapping between Y and x shown in Fig. 2.8b evidently appears to be the kind of non-linearity between "acoustics and perception" which Elman (1977) suggests, and the "stable dichotomy" to which Pastore et al. (1976) refer, and also Ade's (1977), "Type 1" effect.

---

[8] To continue the analogy with optical dispersion, dispersion may result from physical properties of the medium (e.g., refraction by a prism), or as a result of the properties of a particular device (e.g., a diffraction grating).

2.2.2 Dispersion and AX Discrimination

The basic SDT model of discrimination has already been given (Equation 2-11). To incorporate dispersion, the integration is carried out using y as the variable of integration, where y and x are related by Equation 2-13. Once D(x) is specified, the corresponding model for AX discrimination can be computed. The class of dispersion functions of interest at the moment are unimodal and for illustration purposes a Gaussian (Fig. 2.8a) is convenient:

$$D(x) = \frac{1}{2\pi \; \sigma_D} e^{-\frac{(x-x_o)^2}{2\sigma_D^2}} \qquad (2-14)$$

Using Equations 2-11, 2-13 and 2-14 the AX discrimination function can be calculated for all values of x_1 and x_2. The results for various observer criteria, Δy_c, and dispersion widths, σ_D, are shown in Fig. 2.9. Comparing these surfaces with Fig. 2.4 a general similarity is seen, particularly for

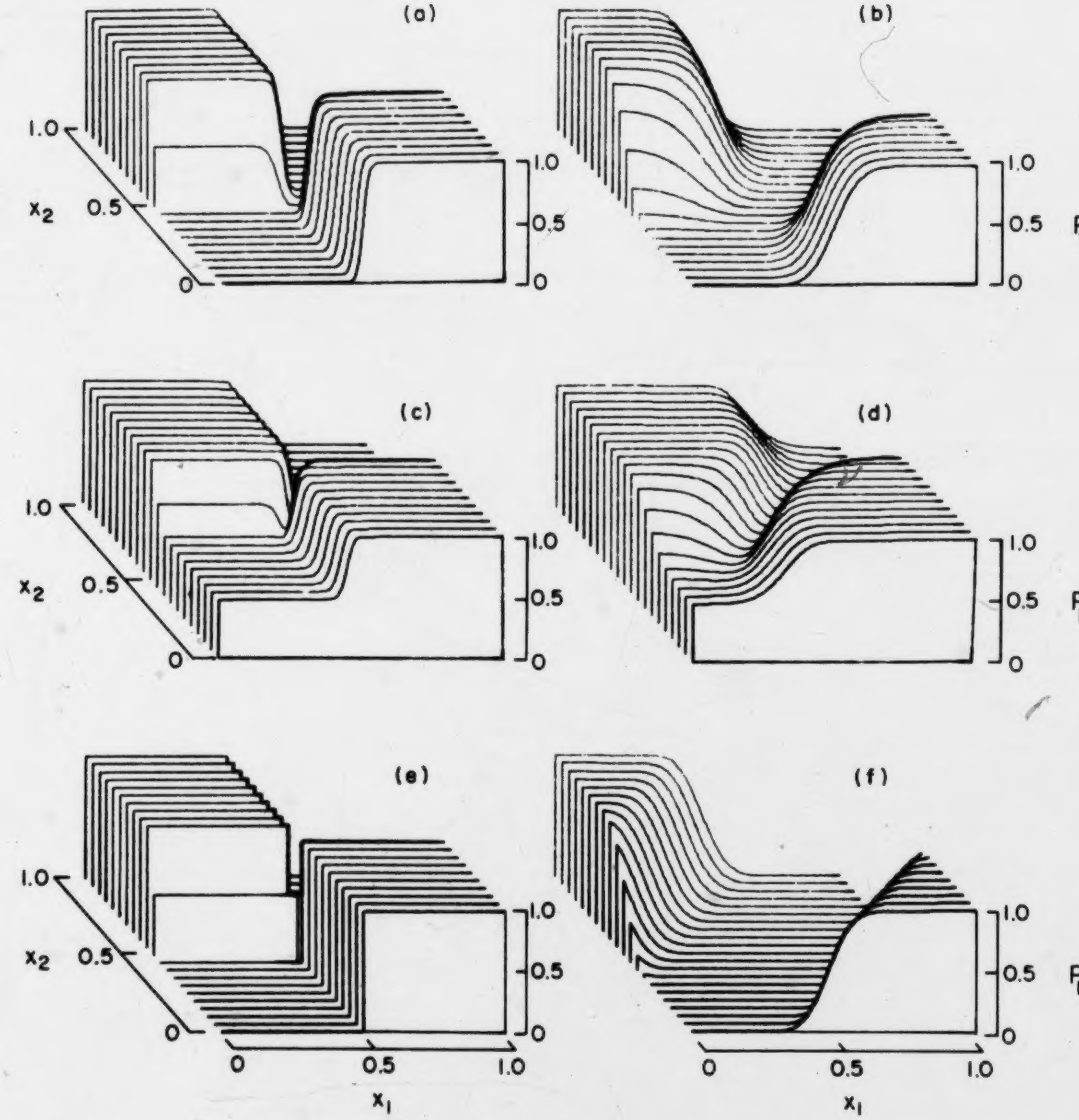


Fig. 2.9. SDT model of AX discrimination, assuming an underlying Gaussian dispersion function. σ is the same for all six figures (σ =0.1). (a) Δy_c =0.5, σ_D= 0. (b) Δy_c=0.5, σ_D= 0.2. (c) Δy_c=0.1, σ_D = 0.05. (d) Δy_c=0.1, σ_D = 0.2. (e) perfectly categorical discrimination (small σ_D). (f) continuous discrimination (large σ_D)

high dispersion (characterized by a small value of σ_D). As is decreased, D(x) approaches a delta function and perception becomes perfectly categorical, as shown in Fig. 2.9e. While this is not physically plausible, it is certainly the desired limiting behaviour of the model. In the other extreme, as the width of the dispersion function increases, D(x) becomes approximately constant over the range of x, and perception becomes "continuous" (compare Fig. 2.9f with Fig. 2.7).

The behaviour of this model for zero or small step sizes is quite different from that for the phonetic memory model. In the present case (see Fig. 2.10), for $x_1 = x_2$, the discrimination function has a constant value (i.e., the number of false alarms) dependent only on the observer criterion Δy_c:

$$P_D = 2(1 - \Phi\left(\frac{\Delta y_c}{\sqrt{2}\,\sigma}\right)) \qquad (2\text{-}15)$$

This is true even in the limit of the dispersion function becoming a delta function, in which case the discrimination function has a singularity at $x_1 = x_2 = 0.5$ (see Fig. 2.9e). As the step size is increased, a peak in the discrimination function appears and broadens in much the same fashion as Fig. 2.5.

In a fixed-standard comparison in which the boundary

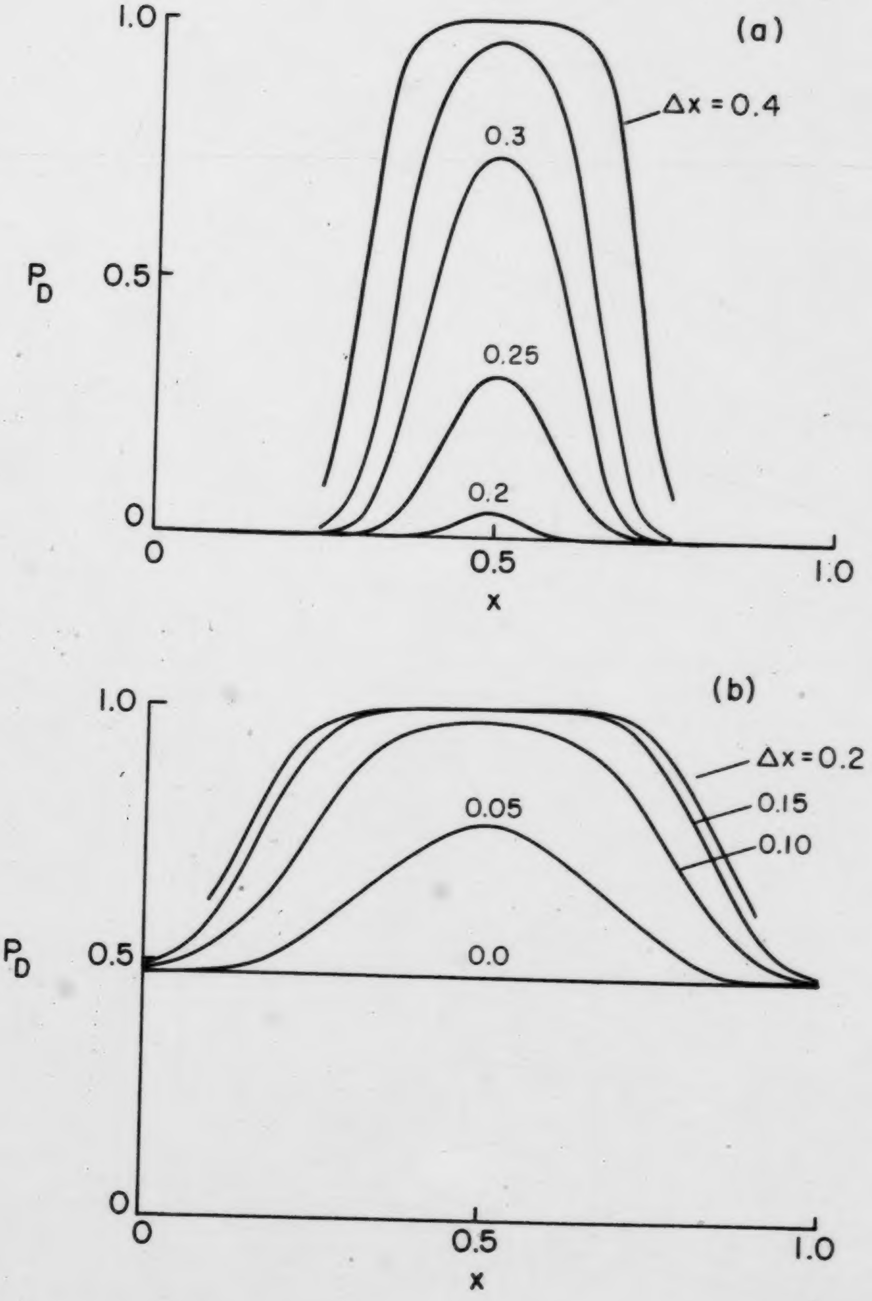


Fig. 2.10. AX discrimination functions for the dispersion model, for various step sizes Δx. (a) $\Delta y_c = 0.5$. (b) $\Delta y_c = 0.05$.

stimulus is chosen as the standard, the discrimination function shows a dip at the boundary, and increases monotonically with distance away from the boundary. (This is to be contrasted with the behaviour of the phonetic-memory AX model which predicted that the discrimination would be constant). This is intuitively satisfying, since a boundary stimulus is a boundary stimulus not because it is an equally good exemplar of either category, but because it is an equally poor exemplar, and hence it should be distinguishable from good exemplars of either category.

It also follows from the phonetic-memory model (Equation 2-5) that when T=0 and the standard stimulus is one of the end-point stimuli (i.e., $p_2=0$) the discrimination function is identical to the identification function. The present model predicts this behaviour only when the width of the dispersion function is sufficiently narrow, and the criterion Δy_c is large (see Fig. 2.11). For small Δy_c, the number of false alarms increases, and the inflection point of the discrimination curve shifts towards the position of the standard stimulus. When Δy_c is made very large, the number of "different" responses in the catagory opposite from that of the standard decreases uniformly. This is to be expected, since a large Δy_c corresponds to a subject responding "same" when the AX stimuli are perceived in clearly opposite phonetic categories. If little or no discrimination within categories is possible, then on the

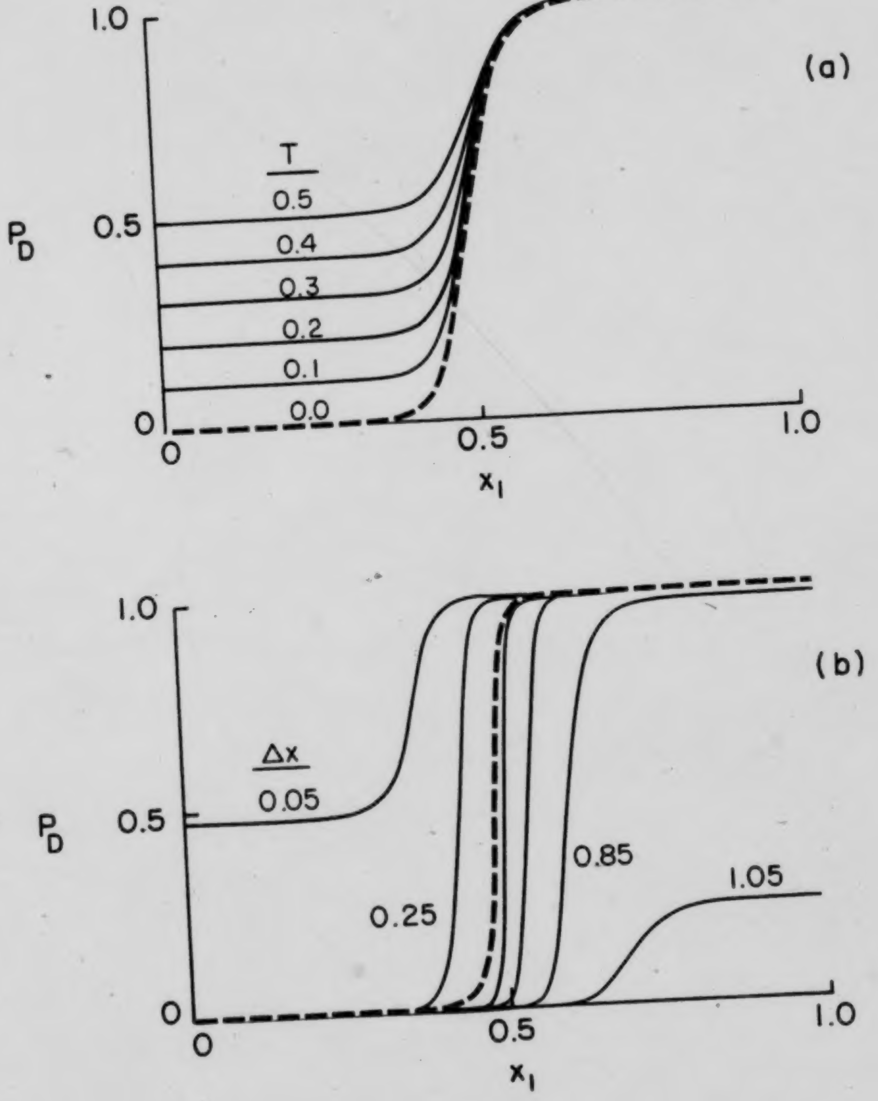


Fig. 2.11. Comparison of variable step size AX discrimination curves when the standard stimulus is $x_2 = 0$. (a) phonetic memory model (Equation 2-5). The various curves are differentiated by different values of T. (b) dispersion model (Equations 2-11, 2-13 and 2-14). The parameters of the curves are Δy_c, the observer criterion. In both (a) and (b), σ is 0.05. The dashed line represents the corresponding identification function (i.e., Equation 2-4)

average he may respond "same" to any stimulus drawn from this category. The degree of categorization of the continuum is indicated by the extent to which the discrimination curve can be "pushed" past the identification curve.

In summary, this "auditory-dispersion" model of the discrimination process predicts the following effects:

(a) for a zero step size, discrimination will be constant at a value dependent on the criterion Δy_c of the observer

(b) for a unimodal dispersion function D(x) the discrimination function for the constant step size paradigm is also unimodal

(c) the number of "different" judgements for within-category stimuli is dependent both on the observer criterion and within-category discriminability

(d) for a fixed-standard test, when an endpoint stimulus is used as the standard, the resulting discrimination curve will be ogival in shape and shifted from the labelling boundary by an amount determined by the observer same-different criterion, Δy_c (as shown in Fig. 2-11b)

(e) a unimodal dispersion function will result in a "categorization" of the continuum, and the width of the transition region of the identification function will reflect both the variance of the internal noise associated with the signal transformation (Equation -9), and the width of the underlying dispersion function

The dispersion AX model shows quite different predictions with respect to the number of false alarms, the number of "different" judgements which result when the stimuli are physically identical. The number of false alarms depends on the observer criterion Δy_c and is given by Equation 2-15 above. However, this condition can result if D(x) becomes zero, since y_1 and y_2 corresponding to two physically different stimuli x_1 and x_2 will be equal. Thus, to the observer, there is no difference between two stimuli which are physically identical and two stimuli which are physically different as long as they map in both cases onto the same value of y. It follows that for a fixed criterion Δy_c,

$$P_D(x_1, x_2) \geqslant P_D(x_1, x_1) \qquad (2\text{-}16)$$

for arbitrary stimuli x_1 and x_2. This is equivalent to saying that two stimuli cannot be more perceptually similar than when they are physically identical.

2.2.3 Finding a Dispersion Function

The primary deficiency of the above model is the ad hoc specification of the dispersion function. Since it cannot be directly observed, it must be inferred by fitting the data to a model in which some explicit form of D(x) is assumed. A preferable approach is to arrive at a theoretical equation for the dispersion curve, in which case Equations 2-11 and 2-13 apply directly. Failing that, a reasonable alternative is to choose some function which has the desired attributes, and use the model to extract estimates of the parameters. This results in a curve-fitting model, but one which at least will allow parameterization of the perceptual continuum. A Gaussian may serve as a suitably flexible choice for a first approximation, but there is no theoretical motivation for this function. In any event, given a dispersion function with width σ_D[9], the degree of categorization of a continuum can be related to the dispersion power of the system, and a suitable index would be[10]

$$\varepsilon = \frac{x_b - x_a}{\beta} \tag{2-17}$$

where x_a and x_b represent the endpoint stimuli and β is the width of the dispersion curve at half height. Continuous perception is then represented by $\varepsilon = 0$ and perfect categorical perception by $\varepsilon = \infty$.

2.3 A DETECTOR MODEL OF CATEGORICAL PERCEPTION

It is a common proposal in speech perception studies that decoding of the speech signal is mediated by acoustic and/or phonetic "feature detectors". The major support for this theory comes from selective adaptation studies, where it is proposed that shifts in the labelling curves observed after repeated presentation of a stimulus results from a desensitization or fatiguing of one of the detectors which span the stimulus continuum (e.g., Eimas and Corbit, 1973;

---

[9] Since the dispersion function is not necessarily Gaussian, it is better to use the full width at half height (i.e., width at half-height) as a measure of the width of the dispersion peak.

[10] Fujisaki and Kawashima (1970) use the index $\Delta x/\sigma_D$ to characterize the fixed step size ABX discrimination curve, but since this involves the step size Δx, it is a measure of the paradigm and not of the continuum.

Miller, 1975; Ainsworth, 1977). The general results of the many selective adaptation studies have not required rejection of this view, although it is not universally accepted that this is the appropriate explanation for the boundary shifts (Simon and Studdert-Kennedy, 1978; Elman, 1977). Part of the difficulty of using the detector construct as a basis for the theory of selective adaptation is its lack of specificity. With the exception of Elman (1977), few computations have been performed to date to investigate what properties such a pair of detectors might have. Elman, investigating the possibility that the observed phonetic boundary shifts could be accounted for by changes in observer criterion, proposes the following detector model:

(a) two detectors span the physical stimulus continuum

(b) the detector response functions are Gaussian

(c) the outputs of the detectors, u_1 and u_2, are compared at a higher level

(d) the phoneme boundary is defined by $u_1 = u_2$

In this section, an analysis of such a two-detector configuration is undertaken. Consider two detectors with response functions given by

$$u_1 = h_1(y) \tag{2-18a}$$

$$u_2 = h_2(y) \tag{2-18b}$$

where y is the perceptual dimension corresponding to the physical continuum x. The detector outputs, U_1 and U_2, will be assumed to be normally distributed random variables with equal variance σ (as in Equation 2-9).[11] Given a transformation $y=g(x)$ between the physical control variable x (the physical parameter which defines the continuum) and y, the perceptual continuum which the detectors span, the detector outputs are

$$u_1 = h_1(g(x)) = H_1(x) \tag{2-19a}$$

and

$$u_2 = h_2(g(x)) = H_2(x) \tag{2-19b}$$

For the present investigation, it will be assumed that the trivial relation y=x holds (i.e., dispersion is constant), in which case $u = h(x)$. Assuming, as does Elman (1977, 1979), that $h_1(x)$ and $h_2(x)$ are Gaussian[12],

---

[11] The assumption of equal variance is not necessary, but in the interests of mathematical tractability, it will be assumed.

$$u_1(x) = \frac{1}{\sqrt{2\pi}\,\sigma_D} e^{-\frac{(x-x_{10})^2}{2\sigma_D{}^2}} \qquad (2\text{-}20a)$$

$$u_2(x) = \frac{}{\sqrt{2\pi}\sigma_D} e^{-\frac{(x-x_{20})^2}{2\sigma_D{}^2}} \qquad (2\text{-}20b)$$

where σ_D is the standard deviation of the detector response function and x_{10} and x_{20} are the locations of maximum sensitivity of the response functions. Two Gaussian detector response functions are shown in Fig. 2.12, where the x continuum ranges from 0 to 1.[13]

As a graphical construct for analyzing the behaviour of such a system, the variables U_1 and U_2 for a given x will be

---

[12] The analysis which follows does not require that the detector functions be Gaussian or even unimodal. However, for computational purposes, some specific form is required and a Gaussian is a convenient choice. Inasmuch as Elman (1977) performed his computations using Gaussian functions, and suggestions have been made that the response functions are possibly Gaussian (e.g., Hanson, 1977), there is neither theoretical motivation nor empirical support for this choice of function. Its only virtue is that it is familiar, unimodal and specified by only two parameters.

[13] Again, since it is assumed that y=x, the response functions can be calculated as if they were spanning the x continuum directly.

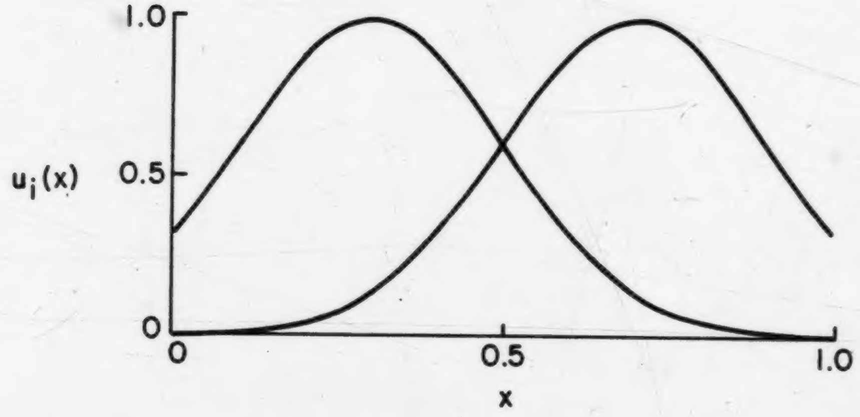


Fig. 2.12. Two Gaussian detector response functions spanning the x-continuum.

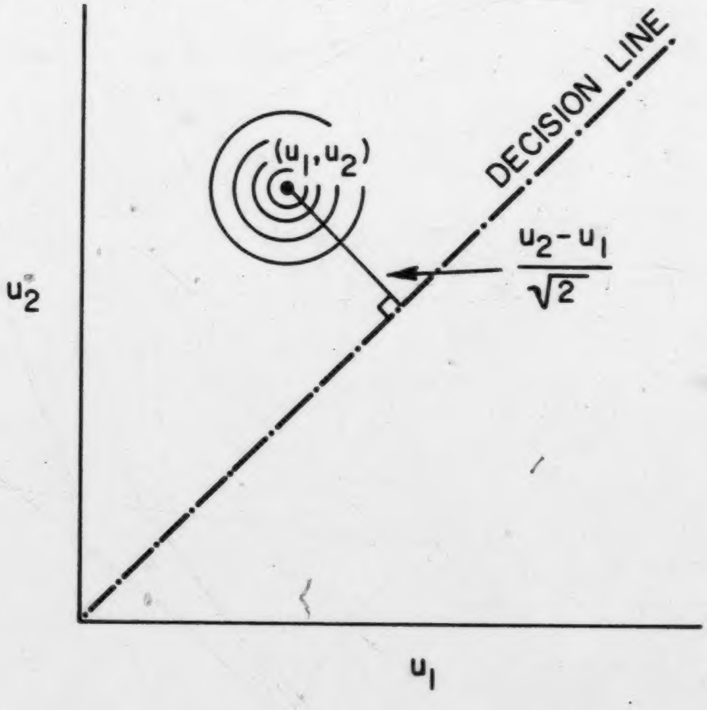


Fig. 2.13. Outputs u_1 and u_2 of the two detectors plotted as a probability density in a two-dimensional decision plane. $u_1 = E(U_1)$ and $u_2 = E(U_2)$ where U_1 and U_2 are independent random variables

displayed as a circular normal probability density function[14] in a u_1 - u_2 signal space (see Fig. 2.13). The presentation of a given stimulus x results in detector outputs u_1 and u_2 as given by Equations 2-20 and, as x takes on values at various points on the x continuum, the points (u_1, u_2) trace out a line in this space. The locus of the points (u_1, u_2) will be referred to as the "stimulus trajectory", and represents a mapping of the x-continuum onto this space. Fig. 2.14 shows the stimulus trajectory for various values of x (in multiples of 0.05) for the Gaussian detector functions shown in Fig. 2.12. The points $(u\ (x), u\ (x))$ are the centroids of a circular normal probability density function of variance σ . The decision line is represented by the dashed line at 45 degrees which corresponds to $u_1 = u_2$. The probability that a given stimulus x will be classified as belonging to category 2 (i.e., $u_2 > u_1$) is then

$$p(u_2 > u_1) = \Phi\left[\frac{u_2 - u_1}{\sqrt{2}\ \sigma}\right] \tag{2-21}$$

where $(u_2 - u_1)/\sqrt{2}$ is the perpendicular distance from point

---

[14] This assumes that the noise sources for the two detectors are uncorrelated. If the noise sources are correlated, the probability density function is still bivariate but not circular. This assumption is made for mathematical convenience but in practice needs to be demonstrated.

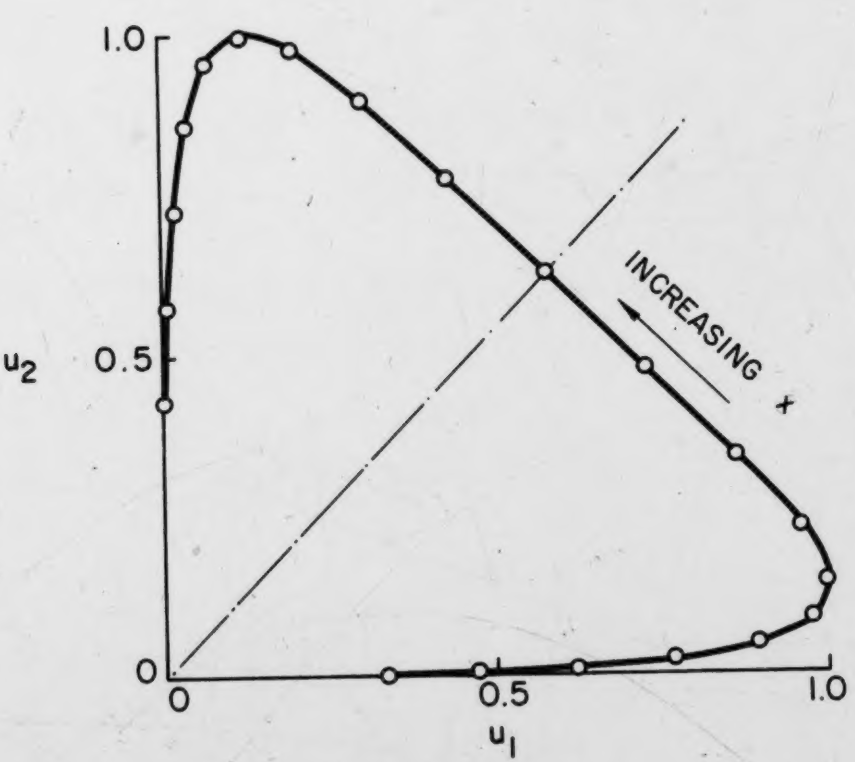


Fig. 2.14. The stimulus trajectory created by the locus of points $(u_1(x),u_2(x))$

(u_1, u_2) to the decision line. The identification function for the stimulus trajectory shown in Fig. 2.14 will be ogival in shape since the decision line is crossed only once. For more complicated detector response functions, as x takes on successive values along the continuum, depending on the nature of the functions $h_1(x)$ and $h_2(x)$, the stimulus trajectory may cross the decision axis only once, or perhaps several times. Multimodal response functions will in general lead to multimodal identification functions.

2.3.1 The Dispersion Function

In order to derive an expression for the dispersion function D(x), it is necessary to define a single decision variable, W, from the two detector outputs. The most obvious choice is

$$W = \frac{U_2 - U_1}{\sqrt{2}} \qquad (2\text{-}22)$$

which is the decision variable used in the identification model. From the previous definition of the one-dimensional dispersion function (Equation 2-13), the corresponding two-dimensional function can be stated:

$$w_{AB} = \int_{C_{\overrightarrow{AB}}} D(\underline{r}) \cdot d\underline{r} \qquad (2\text{-}23)$$

where $\underline{r} = u_1(x)\ \underline{i} + u_2(x)\ \underline{j}$ and A and B are two arbitrary points in the plane connected by the direct arc $C_{\overrightarrow{AB}}$. Now since $d\underline{r} = du_1\ \underline{i} + du_2\ \underline{j}$, and $du_1 = (du_1/dx)\,dx$ etc., Equation 2-23 can be written as

$$w(x) = \int_{x_o}^{x} D(r) \cdot (u_1'\ dx\ \underline{i} + u_2'\ dx\ \underline{j}) \qquad (2\text{-}24)$$

where $u_1' = du_1/dx$ and $u_2' = du_2/dx$. In order for this integral to be path independent (which is equivalent to stating that the similarity between two stimuli x_1 and x_2 depend only on their respective positions in the u_1-u_2 plane), $D(\underline{r})$ must be related to w as follows:

$$D(\underline{r}) = \text{grad } w \qquad (2\text{-}25)$$

The above integral thus becomes

$$w(x) = \int_{x}^{x_o} \left(\frac{\partial w}{\partial u_1} u_1' + \frac{\partial w}{\partial u_2} u_2'\right) dx \qquad (2\text{-}26)$$

and it follows that the dispersion function can be specified as a function of x as:

$$D(x) = \frac{\partial w}{\partial u_1} u_1' + \frac{\partial w}{\partial u_2} u_2' \qquad (2\text{-}27)$$

Now, assuming $w = (u_2 - u_1)/\sqrt{2}$

$$\frac{\partial w}{\partial u_1} = -\frac{1}{\sqrt{2}} \quad \text{and} \quad \frac{\partial w}{\partial u_2} = \frac{1}{\sqrt{2}} \qquad (2\text{-}28)$$

so the dispersion function becomes

$$D(x) = \frac{u_2(x) - u_1(x)}{\sqrt{2}} \qquad (2\text{-}29)$$

which is evidently the form suggested by Elman (1977).[15] For the detector functions defined by Equations 2-20 this dispersion function can be calculated, and is shown as the dashed line in Fig. 2.15. The point of maximum dispersion is located at the category boundary ($u_2 = u_1$).

This is not the only possible choice of decision variable which satisfies relation 2-25. If w is defined as the angle between the line joining the origin and the point (u_1, u_2) and the decision line, i.e.,

$$\tan w = \frac{u_2 - u_1}{u_2 + u_1} \qquad (2\text{-}30)$$

---

[15] Elman does not use the term "dispersion", but uses the phrase "...discriminability, as measured by the difference in slopes" (p. 5). Evidently, he is employing much the same concept but no mathematical details are given.

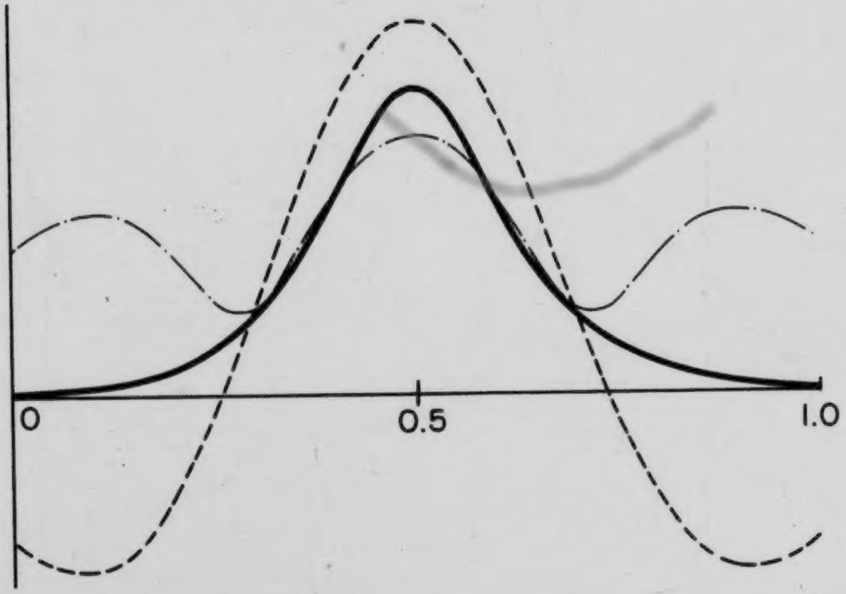


Fig. 2.15. Various dispersion functions for two Gaussian detectors.
(– – – –) Equation 2-29; (———) Equation 2-31;
(—·—) Equation 2-34

the dispersion function is found to be

$$D(x) = \frac{u_1 u_2' - u_1' u_2}{u_1^{\,2} + u_2^{\,2}} \qquad (2\text{-}31)$$

The denominator is just the squared length of the vector $\underline{r} = u_1 \underline{i} + u_2 \underline{j}$, and thus represents a form of "intensity" normalization. The numerator, on the other hand, is just the Wronskian of the two detector response functions $u_1(x)$ and $u_2(x)$, i.e.,

$$W(u_1, u_2, x) = u_1 u_2' - u_1' u_2 \qquad (2\text{-}32)$$

The two functions $u_1(x)$ and $u_2(x)$ are independent only where the Wronskian is non-zero. Thus, assuming that u_1 and u_2 do not go to zero simultaneously, it follows that when the Wronskian is non-zero, the dispersion is non-zero. The solid line in Fig. 2.15 shows this dispersion for two Gaussian detectors.

A third possible measure of similarity of signals in the u_1-u_2 plane is Euclidean distance, measured from some arbitrary but fixed reference point (u_{10}, u_{20}) :

$$w = \sqrt{(u_1 - u_{10})^2 + (u_2 - u_{20})^2} \qquad (2\text{-}33)$$

To express this distance strictly as a function of x, it is necessary to assume some reference point (u_{10}, u_{20}). It is neither clear what this point should be, nor is w independent of this choice. The dispersion function

$$D(x) = \sqrt{1 + \left(\frac{du_2}{du_1}\right)^2} \qquad (2\text{-}34)$$

defines w as distance measured along the direct arc $\underline{r}(x)$ from some arbitrary reference point, but this integral is not path independent. This dispersion function is shown as the dashed-dotted line in Fig. 2.15.

There is no obvious choice between the above three possible dispersion functions. All show a peak at the category boundary, and decrease monotonically away from the boundary, at least in the immediate vicinity of the boundary. Equation 2-31 has the most mathematically desirable properties and is unimodal. This function corresponds to an angular metric. Equation 2-29 has the virtue of using the same decision variable as the labelling function, but it is not unimodal. The last metric, Euclidean distance, was the choice of metric of Zinnes and Wolfe (1977) in their formulation of a model of same-different discrimination for a two-dimensional visual task. However, the w-x mapping given by Equation 2-33 is dependent on the particular path $\underline{r}(x) = u_1(x)\ \underline{i} + u_2(x)\ \underline{j}$, and hence is a more a property of the detector functions than it is of the u_1-u_2 space. For this reason, it will not be considered further.

2.3.2 Calculating the AX Discrimination Function

Given the dispersion function D(x) defined by Equation 2-30, the AX discrimination function can be calculated from Equations 2-11 and 2-13. For the case of two Gaussian detectors, the corresponding AX discrimination functions are shown in Fig. 2.16. It can be seen that this AX discrimination function is virtually identical to that shown in Fig. 2.9, and thus has similar properties.

In the spirit of the concept of "detector", the variables u_1 and u_2 used in the above analyses presumably represent some form of neural excitation associated with the detection of signals specified by the x continuum. The closer the stimulus is to the detector maximum, the greater the degree of excitation. If this is the case, the output of the detector (for a phonetic continuum) is the "phonetic value" of the stimulus. The recognition of a particular phonetic category is then dependent on the relative strengths of excitation of two neural populations which correspond to the two detectors.

Discrimination, then, presents an interesting problem. If perception is mediated by detectors in the form described above, then in order for discrimination to occur at a sub-

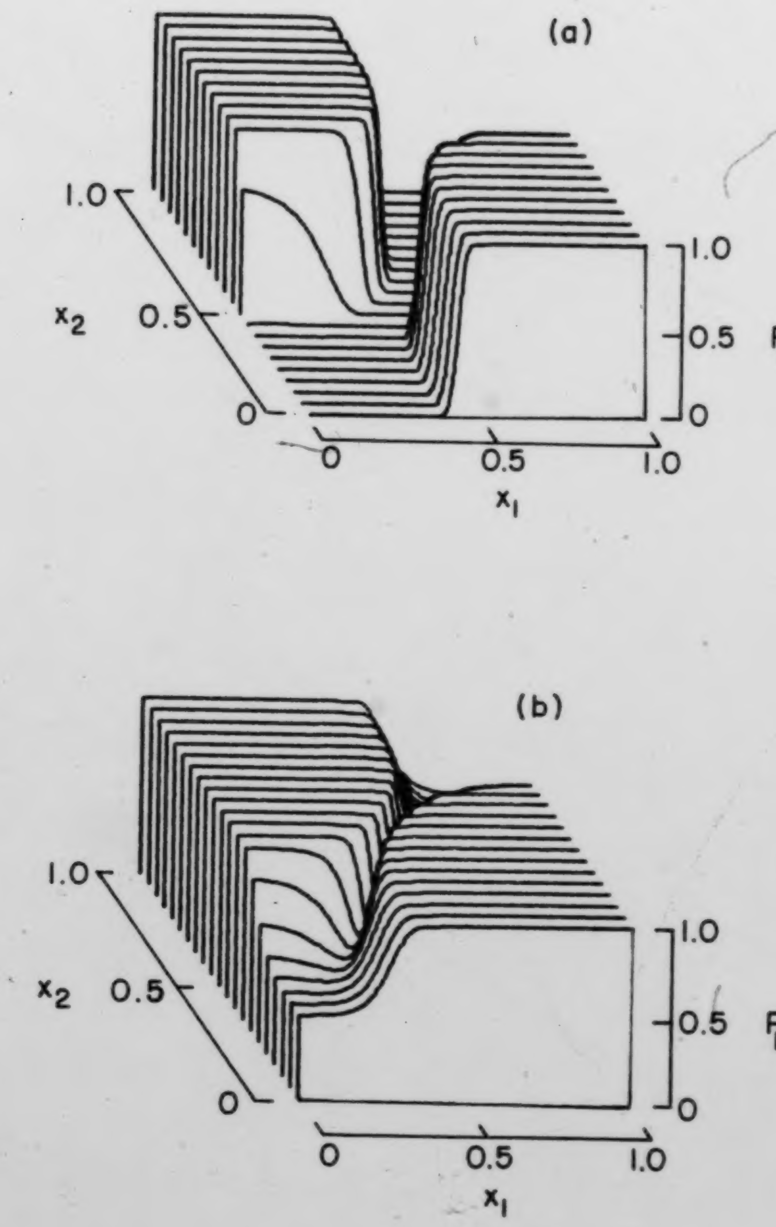


Fig. 2.16. Calculated AX discrimination functions for two Gaussian detector response functions positioned at x_{10} = 0.3 and x_{20} = 0.7 (i.e., Fig. 3.12). The dispersion function is given by Equation 2-31. (a) Δy_c = 0.5, (b) Δy_c = 0.05, where Δy_c is the criterion for the same/different decision.

phonetic level, the outputs of both detectors must leave traces. That is, the values of both u_1 and u_2 must be "remembered". For phonetic level discrimination, on the other hand, only which detector was excited the most strongly need be remembered. An alternative possibility is that the auditory representation of the stimulus (i.e., the value of x) is remembered. In general, if the mapping $y = g(x)$ shows no enhanced dispersion at any point along the continuum, then discrimination will not be enhanced, as shown in Fig. 2.7. If y does show enhanced dispersion, then a "natural perceptual boundary" may exist along the x continuum. It is conceivable that speech categories would be structured around any such extant natural perceptual boundaries rather than the converse.

2.4 APPLICATION TO SELECTIVE ADAPTATION

The model of categorical perception as mediated by a two-detector configuration will now be investigated with respect to selective adaptation. Since the effect of adaptation, as commonly supposed, is to desensitize one of the detectors, this can be modelled by incorporating scaling factors a_1 and a_2 into the detector response functions H_1 and H_2 defined by Equations 2-20. That is,

$$u_1(x) = \frac{a_1}{\sqrt{2\pi}\,\sigma_D} e^{-\frac{(x-x_{10})^2}{2\sigma_D^2}} \qquad (2\text{-}35a)$$

$$u_2(x) = \frac{a_2}{\sqrt{2\pi}\,\sigma_D} e^{-\frac{(x-x_{20})^2}{2\sigma_D^2}} \quad (2\text{-}35b)$$

The boundary stimulus is characterized by the value of x for which $u_1(x) = u_2(x)$, in which case

$$a_1 e^{-\frac{(x-x_{10})^2}{2\sigma_D^2}} = a_2 e^{-\frac{(x-x_{20})^2}{2\sigma_D^2}} \quad (2\text{-}36)$$

The solution for x is

$$x = \frac{x_{10} + x_{20}}{2} - \frac{\sigma_D^2 \ln \frac{a_2}{a_1}}{x_{20} - x_{10}} \quad (2\text{-}37)$$

Now, since the unadapted boundary is just $x_b = (x_{10}+x_{20})/2$, Equation 2-37 becomes

$$x_s = -\frac{\sigma_D^2 \ln \frac{a_2}{a_1}}{x_{20} - x_{10}} \quad (2\text{-}38)$$

where $x_s = x - x_b$ is the boundary shift. This has three implications: first, desensitization of one of the detectors will cause a boundary shift, and second, desensitization of both detectors simultaneously and by the same fraction (i.e., such that a_2/a_1 does not change) will cause no change in the boundary. The first result apparently has been verified many times in the selective adaptation literature, and the second result has also been demonstrated (Miller, 1977; Sawusch and Pisoni, 1976). The third implication is that, ceteris paribus, larger boundary shifts will occur for less strongly categorized continua (i.e., with larger σ_D).

2.4.1 Discrimination Under Conditions of Adaptation

Cooper (1974) investigated the shift in the peaks of the ABX discrimination curve for /ba/, /da/ and /ga/ adaptors on an F_2-F_3 continuum. His results show that the shift in the peak of the discrimination function is in the same direction and of approximately the same magnitude as the shift in the boundary of the corresponding identification function. This suggests that there is an intimate relationship between the location of the identification boundary and the peak of the discrimination function (or in the light of the previous discussions, the peak of the dispersion function).

Consider the dispersion function defined by Equation 2-31. The generalization of this function for the Gaussian detector functions given by Equation 2-35 is

$$D(x) = \frac{(x_{20} - x_{10})\, u_1 u_2}{\sigma_D\, (u_1^2 + u_2^2)} \qquad (2\text{-}39)$$

Now, the peak of the dispersion function occurs for x such that dD/dx = 0. Differentiating Equation 2-39 and setting it to zero produces

$$(u_1 - u_2)(u_1 + u_2)D(x) = 0 \qquad (2\text{-}40)$$

provided that u_1 and u_2 do not go to zero simultaneously. Since the Dispersion function D(x) (for Gaussian detectors) is non-zero for finite values of x, it follows from Equation 2-40 that

$$u_1 = u_2 \qquad (2\text{-}41)$$

Thus, this detector model of categorical perception has the property that the peak of the dispersion function (and hence discrimination function) will always coincide with the phonetic boundary (for detectors with Gaussian response functions).

2.4.2 The Effect of Adaptation on the Stimulus Trajectory

The effect of adaptation on the stimulus trajectory is shown in Fig. 2.17. Desensitizing one detector is equivalent to scaling the corresponding axis of the decision plane by the same factor. This causes a distortion of the stimulus trajectory such that the point $u_1 = u_2$ now corresponds to a different value of x. Two results automatically follow from this. One, a desensitization of one or both detectors is equivalent to a change in response bias. For a desensitized detector configuration, the unbiased decision line is $u_1 = (a_2/a_1)u_2$, which is

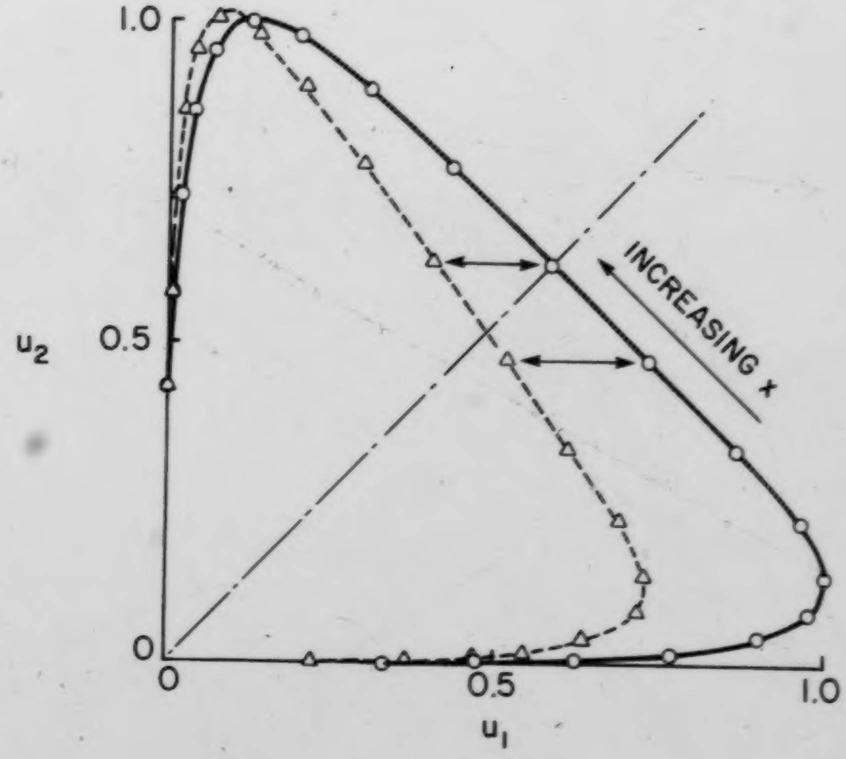


Fig. 2.17. Effect of adaptation on a two-detector system to a stimulus from category 1 (i.e., only u_1 is affected). The arrows connect identical values of x. Note that the category boundary (the point where the stimulus trajectory crosses the decision line) shifts towards the category of the adaptor

equivalent to a decision line at an angle ϕ where $\tan\phi = a_1/a_2$. However, a change in bias also corresponds to a change in the angle of the decision line, e.g., $u_2=bu_1$, where $b=1$ corresponds to the 45 degree decision line. Under conditions of adaptation, the category boundary is then defined by $u_2=(ba_1/a_2)u_1$, which shows that, according to this model, response bias and adaptation are formally inseparable. According to the present model, there is no way to distinguish between the detector desensitization and response bias accounts of phonetic boundary shift by simply measuring the boundary shift.

This model of selective adaptation suggests yet another effect. Assuming that the output of the detectors increases monotonically with stimulus intensity, e.g.,

$$u_1 = h_1(I)u_1(x) \tag{2-42a}$$

$$u_2 = h_2(I)u_2(x) \tag{2-42b}$$

radial distance in the u_1-u_2 plane ought to represent an intensive aspect of the stimuli. If the detector response curves merely reflect the sensitivity of the detector to certain stimuli, then decreasing the intensity of the stimulus should also produce a translation of the stimulus trajectory in the direction of the origin. Thus, in this model, decreasing the sensitivity of a detector is equivalent to decreasing the intensity of the signal to which the detector responds. Miller, Eimas and Root (1977) conducted a selective adaptation experiment using /bae/, /dae/ and /gae/ stimuli each of which was constructed with nine levels of attentuation of F_2 and F_3 with respect to F_1. Their results show that after adaptation to, say, /bae/, to obtain a level of identification equal to the pre-adaptation condition, the /bae/ must be more intense. Similar results were found with the /gae/ adaptor. This is the only experiment to date which has attempted to see if the desensitization of a particular detector ("channel of analysis" in their terminology) can be restored by an equivalent increase in intensity, and their results tend to support the predictions of the above model.

2.4.3 A New Experimental Paradigm

The detector model presented above suggests a new experimental paradigm. Consider a signal which simultaneously contains the acoustic cues corresponding to two positions x_1 and x_2 along the x-continuum.[16] Such a signal, according to this detector model, will generate u_1

---

[16] In Chapter 3 it will be shown how to construct such a signal.

and u_2 as

$$u_1 = u_1(I,x) = h_1(I)u_1(x) \qquad (2\text{-}43a)$$

and

$$u_2 = u_2(I,x) = h_2(I)u_2(x) \qquad (2\text{-}43b)$$

Holding x_1 and x_2 fixed but varying the relative intensities I_1 and I_2 of these two cues gives

$$u_1 = h_1(I_1)u_1(x_1) \qquad (2\text{-}44a)$$

and

$$u_2 = h_2(I_2)u_2(x_2) \qquad (2\text{-}44b)$$

from which it is seen that the outputs of the detectors, u_1 and u_2, can be manipulated by altering intensities I_1 and I_2. If, for instance, I_1 and I_2 are related by

$$I_1 = I_o - I_2 \qquad (2\text{-}45)$$

then if h(I) is monotonic with I and h(0)=0, it follows that there exists a value of I for which $u_1 = u_2$. That is, relative intensity I will define a continuum between the two categories, with $I=0$ representing one category and $I=I_o$ representing the other. A category boundary will exist for the value of I for which $u_1=u_2$. Thus, as I is varied from 0 to I_o, again u_1 and u_2 will trace out a trajectory as shown in Fig. 2.18. Since this stimulus trajectory is similar to that shown in Fig. 2.14, it would appear that, if these assumptions are correct, a "relative intensity continuum" will be formed. As will be shown in Chapter 3, such a continuum, for some combinations of speech sounds, is also categorically perceived.

2.5 SUMMARY

In summary, the analysis of the two-detector model given above shows that a two-detector configuration produces a dispersion of the decision variable $Y = U_1-U_2$ (i.e., the detector outputs) with respect to the physical continuum "x". The shape of the dispersion function is strictly determined by the detector response functions. Two detectors which have opposite slopes at the phonetic boundary will always lead to enhanced dispersion in the vicinity of the category boundary (which is the point where the detector outputs are equal). The degree of dispersion (and hence discrimination) is a strong function of the slopes of the detector response functions near the boundary: the steeper the slope, the larger the dispersion. In discrimination, the subjective task is one of perceiving

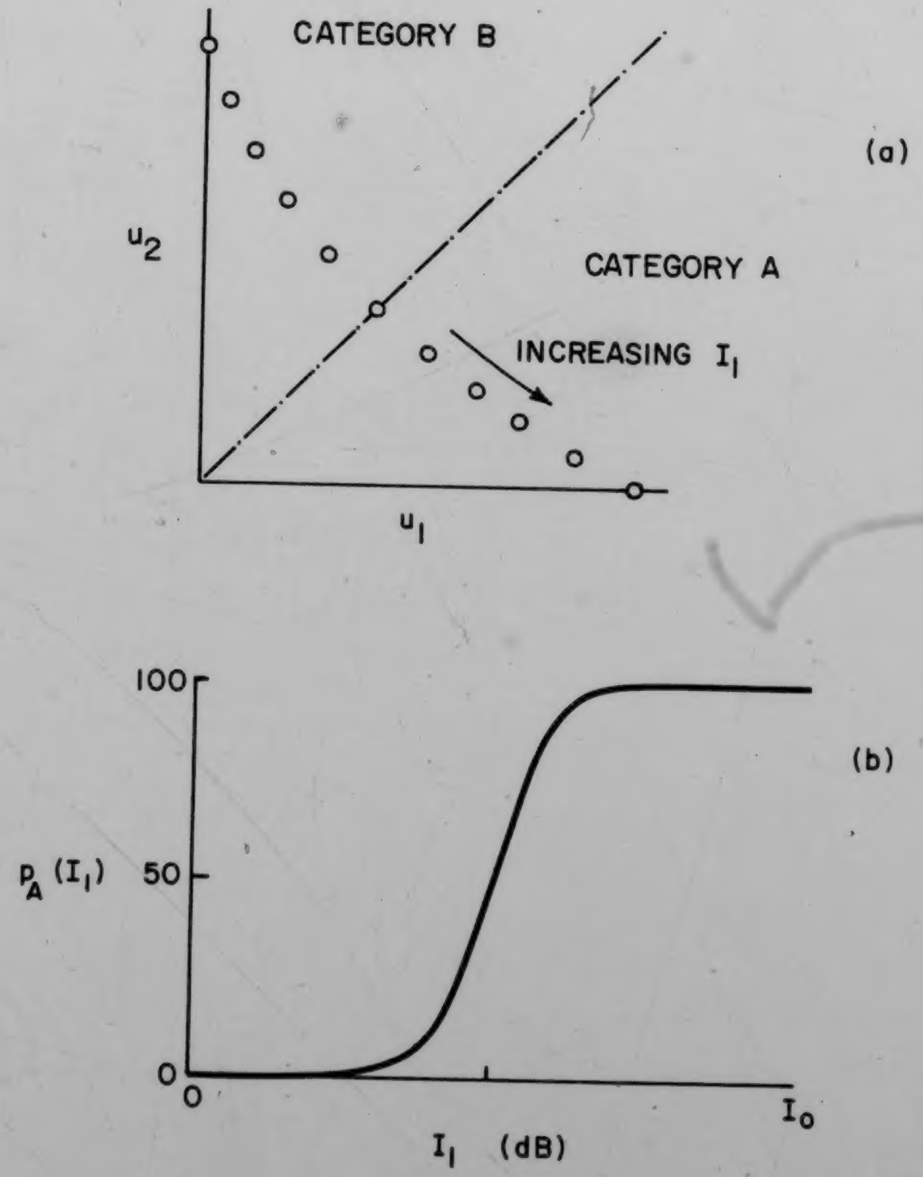


Fig. 2.18. (a) Hypothetical stimulus trajectory for $u_1=h_1(I_1)u_1(x_1)$ and $u_2=h_2(I_0-I_2)u_2(x_2)$ for fixed x_1 and x_2.
(b) corresponding identification curve as I_1 is varied from 0 to I_0

differences in sensation. These differences may be small or large, depending on the physical difference between the signals, and the extent to which the auditory system is sensitive to these differences. It may well be the case that phonetic differences are merely "large acoustic differences". This does not imply that a phonetic level of processing per se does not exist, nor does it belittle the role of phonetic memory in various experimental paradigms. The ABX or oddity paradigms, for instance, without doubt place stringent demands on phonetic memory (Miller and Eimas, 1977; MacMillan et al., 1977; Pisoni and Lazarus, 1974; Pisoni, 1973; Fujisaki and Kawashima, 1970) and the phonetic memory model is likely appropriate for these conditions. The AX paradigm, on the other hand, places considerably less load on phonetic memory, and a model based on auditory rather than phonetic differences may be appropriate to account for observed discrimination data.

Although few specific suggestions have been made for the embodiment of feature detectors in neurophysiological terms (Abbs and Sussman, 1971, suggest the term "neurosensory receptive field"), it follows that in order for the concept to have substance, a specific formulation must be given. Simon and Studdert-Kennedy (1978) caution against a literal interpretation of the detector metaphor, and suggest the use of the term "channel of analysis" as being more neutral. This is sound advice, but the issue is more than one of just terminology. The question is not whether or not there exist special detectors specially tuned to selected attributes of speech signals, but whether or not the nervous system can behave as if this were the case. With a fixed set of stimuli drawn from a continuum, the "detector" metaphor may be a perfectly suitable characterization. The analysis conducted in this chapter is a mathematical characterization of the detector metaphor, and does not make any specific claims regarding the physical existence or physiological makeup of such detectors, were they to exist. Rather, it represents an attempt to quantify proposals which have appeared from time to time in the literature, and the appropriateness of the metaphor may hopefully be clarified by so doing.

CHAPTER 3

CATEGORICAL PERCEPTION OF A RELATIVE INTENSITY CONTINUUM

In this chapter, a new technique for investigating categorical perception is presented. The continuum in this case is the relative intensity of two CV stimuli whose time waveforms are added together. When the two CV (or VC) stimuli contain phonemes which normally occur in the same position in the syllable (i.e., initial or final) and which lie on an acoustic continuum, the percepts fuse. When the relative intensities of the two components (C_1 and C_2) are varied, a "continuum" is created which is categorically perceived. For syllable-initial stop consonants (e.g., /b/ and /d/), the relative intensity continuum has all the properties of an F_2-F_3 continuum, but has the distinct advantage of being specified by a single physical parameter. In this chapter, various experiments are conducted to investigate the origin and generality of the effect. ABX and AX discrimination results show that disciminability is poor except when components C_1 and C_2 are nearly equal in intensity. The dispersion model of AX discrimination (Section 2.2) is fitted to the experimental data, and the fit is observed to be quite satisfactory. On the basis of the observed results, the phonetic-memory model of categorical perception can be rejected.

3.1 CREATION OF THE TEST STIMULI

Fig. 3.1 shows schematically how an ambiguous CV (/bae/-/dae/) signal is created.[1] Two /bae/ and /dae/ signal waveforms with similar f_0's are added together, and the result is a waveform which contains the phonetic cues for both /bae/ and /dae/. Such a signal is found to be perceptually ambiguous, and is perceived as either /bae/ or /dae/. The quality of the resulting stimulus depends on how well the two component stimuli are aligned, and mutual interference can be minimized if the two waveforms are

---

[1] Most of the experiments described in this thesis employed this particular pair of stimuli. For this reason, the stimulus construction procedure will be described in terms of these stimuli. However, the procedure (and the perceptual effect) is quite general, and can be applied to various other signal waveforms, including syllable-final stop consonants.

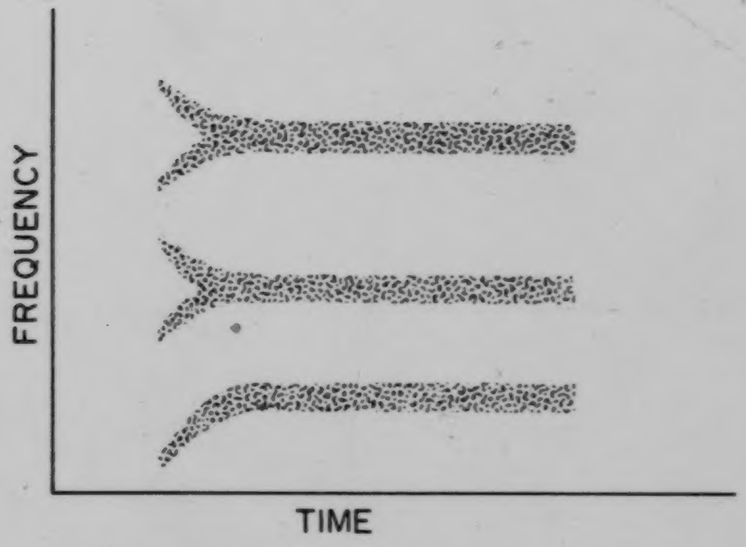


Fig. 3.1. Schematic formant transitions for a composite /bae/-/dae/ stimulus

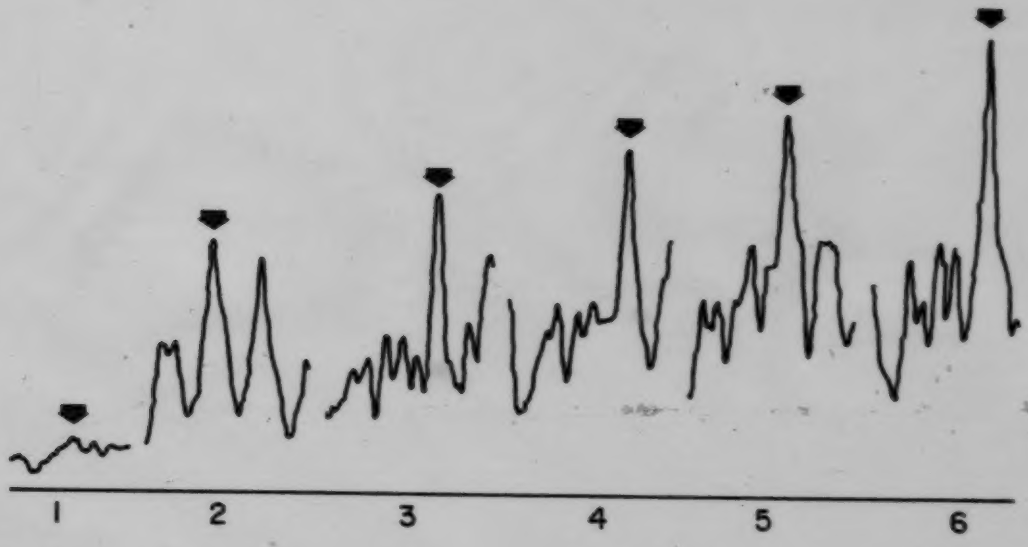


Fig. 3.2. Example of covariance trace for pitch periods one through six for /bae/ and /dae/. The arrows indicate the points of maximum correlation between corresponding /b/ and /d/ pitch periods

suitably aligned.

The procedure for "suitably" aligning the stimuli is best explained by means of an example. Multiple tokens of /bae/ and /dae/ were recorded by the author with approximately equal fundamental frequencies (f_0 = 100 Hz) and steady-state vowel formant values. The recording was carried out in an acoustically isolated chamber using a TEAC AR-70 tape recorder. A /bae/-/dae/ pair was then selected on the basis of judgements of the similarity of the f_0's and the compatibility of the formant values of steady state vowels, as determined from Sonagrams of the stimuli and plots of the signal waveforms. These tokens of /bae/ and /dae/ were digitized at 16 kHz and stored in a disk file for later processing.[2] The first nine pitch periods of these two waveforms were extracted and stored separately.[3] The waveform with the smallest f_0 (/b/ in this example) was selected as the "standard", to which the other signal (/d/) was then temporally aligned. The alignment procedure was as follows: separate signals $/b/_j$ and $/d/_j$ (j=1,2,...,9) were created from each of the pitch periods of the digitized /b/

---

[2] All signal preparation and presentation was carried out using an programming system for the PDP-12 designed by the author (see Stevenson and Stephens, 1978).

[3] The following notation will be used: /bae/ will refer to the entire CV syllable, while /b/ will refer only to the extracted formant transitions from that syllable. /ae/ will represent the steady-state vowel, which is the tenth through last glottal pulse of the CV waveform.

and /d/ waveforms . Each $/d/_j$ was then aligned with its corresponding $/b/_j$ by calculating

$$r_j = \sum_{i=1}^{n_j} s_{b_{ij}} s_{d_{ij}} \qquad (3\text{-}1)$$

where s_b and s_d are the waveform amplitudes of /b/ and /d/ respectively. The two signals were considered aligned for the offset which yielded a maximum r_j. (Fig. 3.2 shows the plots of r_j for the first six pitch periods). After each of the $/d/_j$ was treated in this fashion, a new set of /d/ formant transitions was created by concatenating these adjusted /d/ pulses. The resulting /d/ waveform now matched the /b/ waveform on a pulse for pulse basis (see Fig. 3.3). To complete the stimulus preparation, the /d/ was then scaled so that its overall intensity was equal to that for the /b/, where the intensity was measured by

$$I = \sum_{i=1}^{N} s_i^{\,2} \qquad (3\text{-}2)$$

The summation is taken over the nine pitch periods of the formant transitions (total of N points).

These /b/ and /d/ signals, together with the steady-state vowel /ae/ (which was the tenth through last pitch period of /bae/) were stored in a file, and were used as the

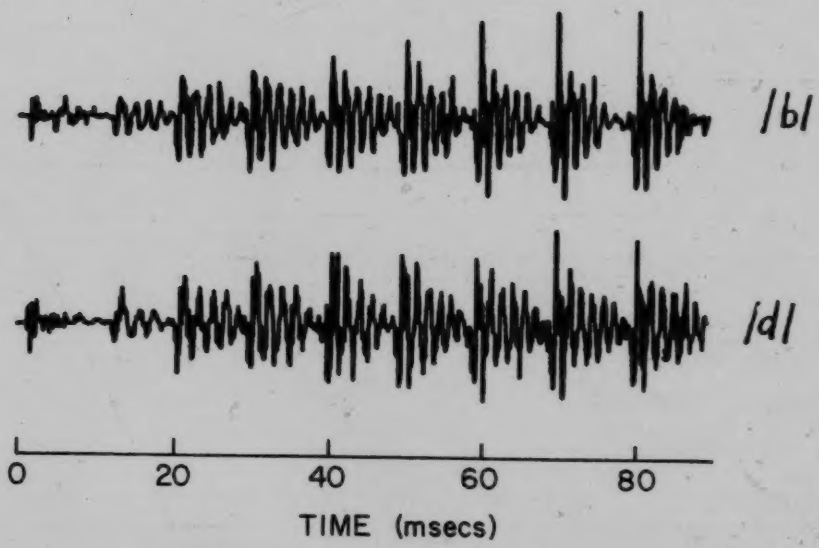


Fig. 3.3. Aligned /b/ and /d/ formant transitions

basis signals for construction of all test stimuli. Presentation stimuli were prepared from these three signals by first loading them into the computer memory and then adding the /b/ and /d/ formant transitions together point by point according to

$$s_i' = \alpha s_{b_i} + (1-\alpha) s_{d_i} \tag{3-3}$$

Finally, the resultant signal, s', was concatenated to the steady-state vowel /ae/. The linear weighting of the /b/ and /d/, along with the convergence of both of these waveforms to the same steady-state vowel ensured continuity of the amplitudes of the mixed formant transitions and steady-state vowel.

Fig. 3.4a shows the relative amplitudes of the /b/ and /d/ waveforms as a function of the weighting parameter α . Because the amplitudes of the signal waveforms were varied linearly, the intensity of the waveforms varied quadratically. Since both signals were of the same duration ($T \sim 90$ milliseconds), their intensities can be represented by the total energies

$$E_b = \sum_{i=1}^{N} s_{b_i}^2 \tag{3-4a}$$

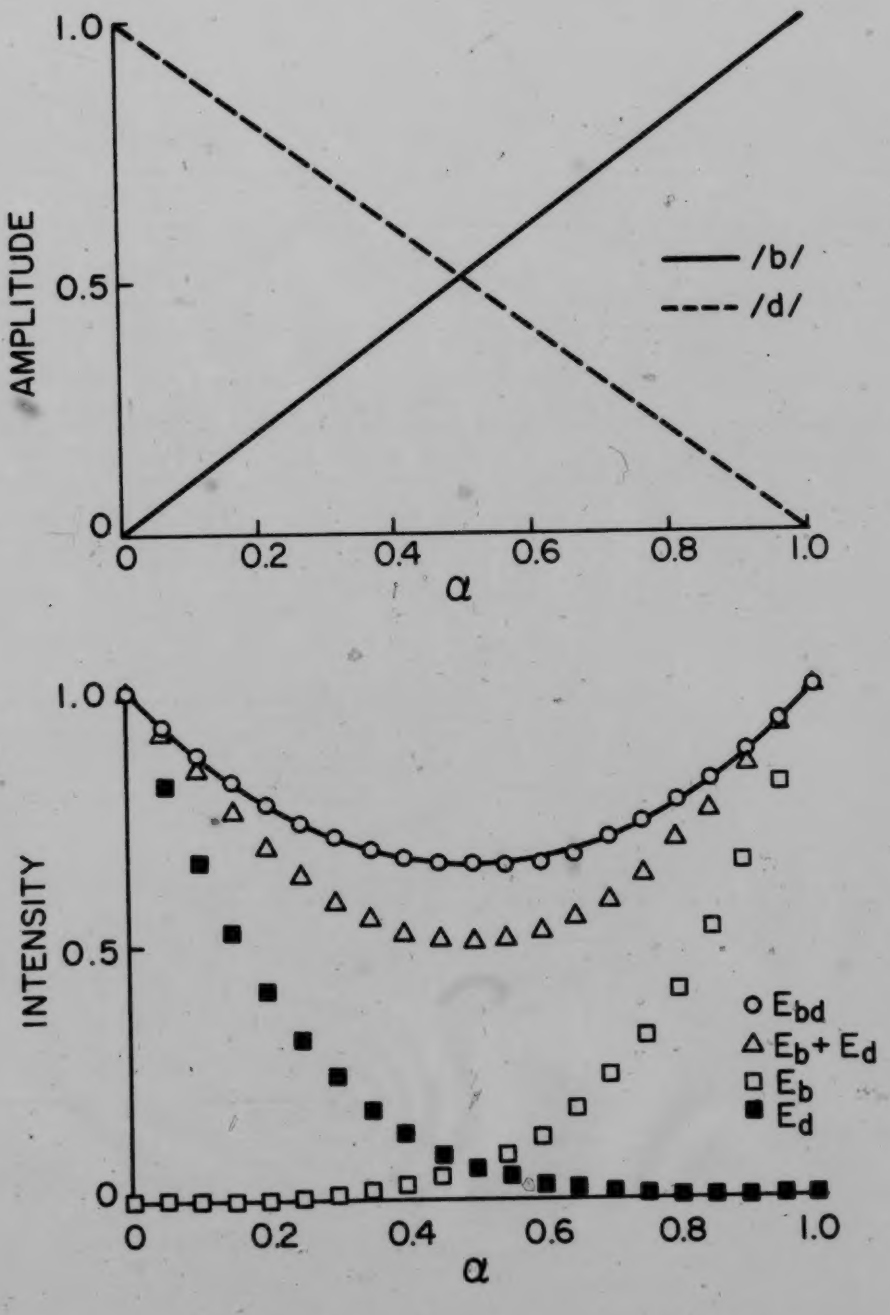


Fig. 3.4. (a) variation of amplitude of /b/ and /d/ waveforms with parameter α. (b) variation of E_b, E_d and E_{bd} with α. The solid line is given by Equation 3-6

and

$$E_d = \sum_{i=1}^{N} s_{d_i}^{\,2} \qquad (3\text{-}4b)$$

where $s_b(t)$ and $s_d(t)$ are the time waveforms of the /b/ and /d/ formant transitions. The intensity of the composite signal can thus be expressed as

$$E_{bd} = \alpha^2 E_b + (1-\alpha)^2 E_d + 2\alpha(1-\alpha)\rho\sqrt{E_b E_d} \qquad (3\text{-}5)$$

where ρ is the Pearson product-moment correlation coefficient. Fig. 3.4b shows the measured E_{bd} as a function of the weighting parameter α. As expected, the curve is quadratic in shape, and has a minimum at $\alpha = 0.5$. Since the /b/ and /d/ formant transitions were equated for overall intensity, the E_{bd} curve reaches the same maximum value at the extreme positions $\alpha = 0$ and $\alpha = 1$. The solid line through the data points in Fig. 3.4b is a quadratic curve fitted by least squares:

$$E = 1.00 - 1.36\,\alpha + 1.36\,\alpha^2 \qquad (3\text{-}6)$$

This equation was used to eliminate differences in overall intensity in the discrimination and binaural experiments described later in this chapter. Letting $E_b = E_d = 1$, Equation 3-5 becomes

$$E_{bd} = 1 - 2(1-\rho)\alpha + 2(1-\rho)\alpha^2 \qquad (3-7)$$

whence $\rho = 0.32$.

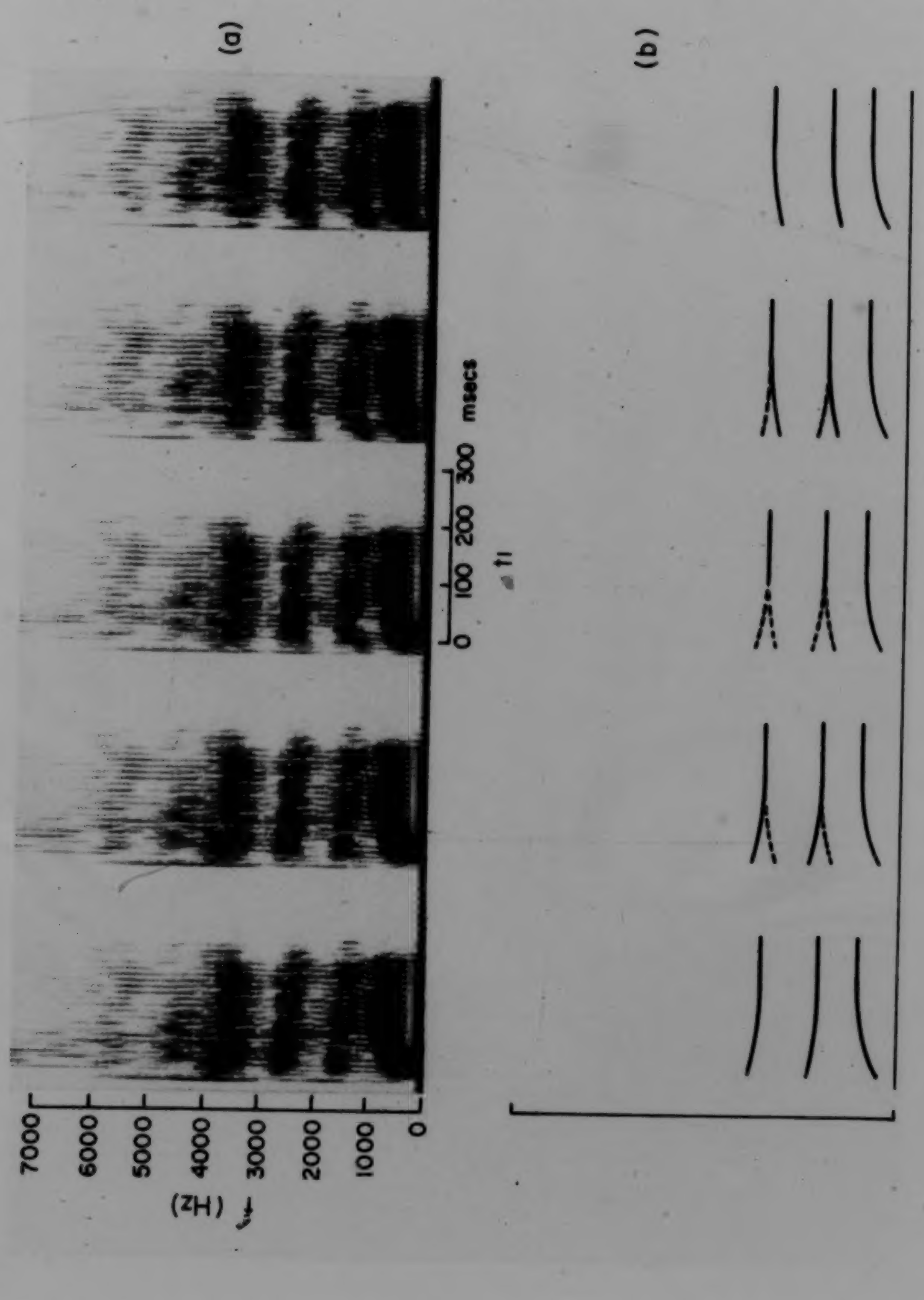
3.2 IDENTIFICATION CURVES

Informal experiments indicated that a sharp transition between phonetic categories existed where the intensities of the two CV components were approximately equal (i.e., $\alpha = 0.5$). A series of experiments was carried out to determine the nature of the identification function for the continuum determined by the parameter α. All stimuli were constructed according to the procedure outlined in Section 3-1 above, and were presented on-line to one or more subjects in a quiet listening environment as detailed below. The stimulus pairs were /bae/-/dae/, /bɛt/-/dɛt/ and /ra/-/la/. Sonagrams of some of the /bae/-/dae/ combinations are shown in Fig. 3.5.

3.2.1 Experimental Setup

The physical arrangement of the computerized presentation facility is shown in Fig. 3.6. The digitized stimuli were converted into analogue voltages at a sampling rate of 16 kHz by 10-bit D/A converters which were part of the PDP-12 configuration. The stimuli were delivered to the remote listening station over lines with a measured 55-60 dB

Fig. 3.5. Sonagrams of /bae/-/dae/ stimuli. From left to right the values of α are: $\alpha=0$ (extreme left, corresponding to a pure /dae/), $\alpha=0.25$, $\alpha=0.5$, $\alpha=0.75$, $\alpha=1.0$ (extreme right, corresponding to a pure /bae/). The line drawing below illustrates the formant composition


(a)
f (Hz)
7000
6000
5000
4000
3000
2000
1000
0
0
100
200
300
msecs
t
(b)

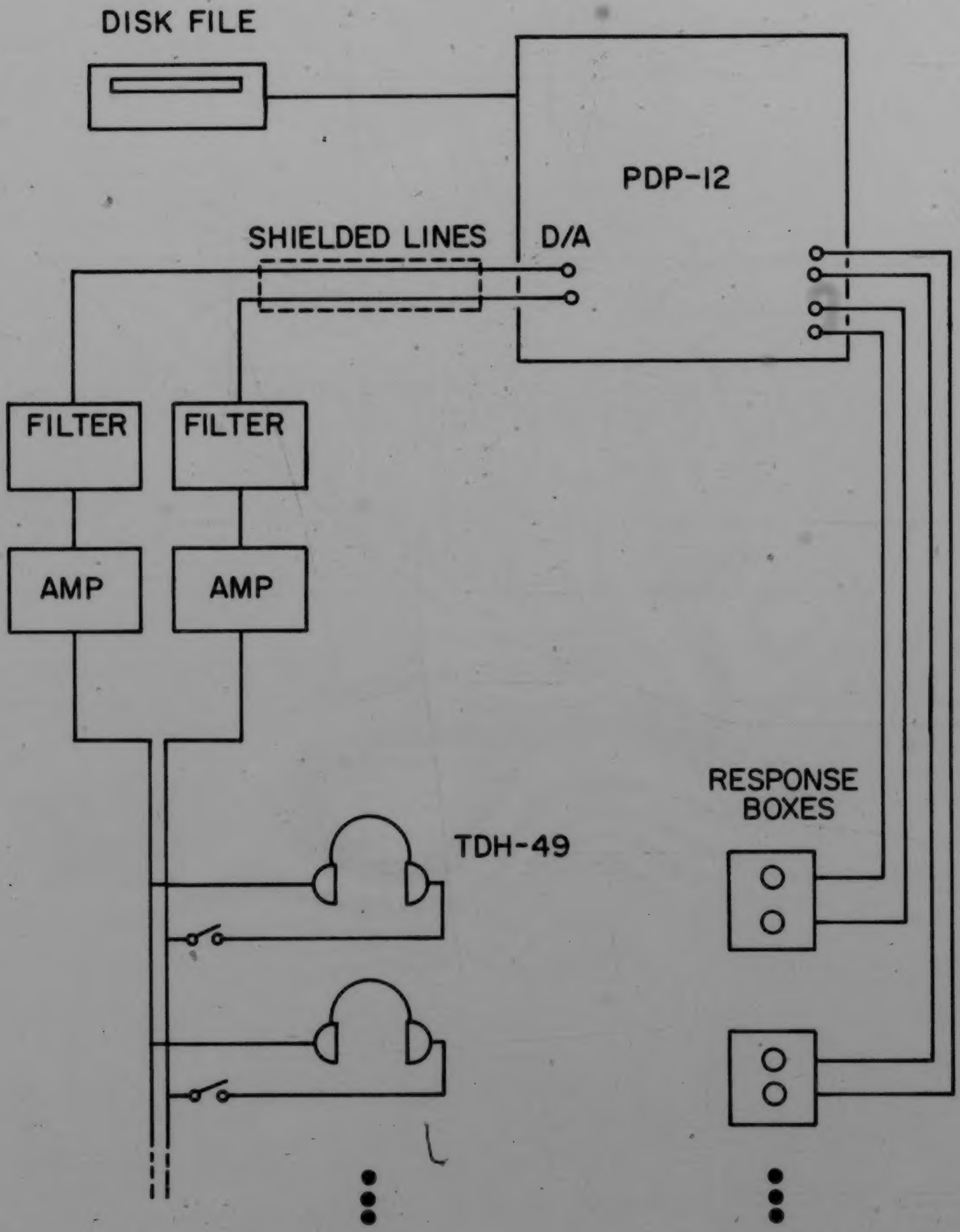


Fig. 3.6. Computerized presentation facility. The headphones and response boxes are in a quiet listening environment isolated from the computer

signal-to-noise ratio and then filtered by a Rockland Series 1520 filter (Butterworth) set to lower and upper cut-off frequencies of 70 and 7000 Hz respectively. The output of the filter was amplified by a Braun amplifier (Type CSV 250) which was linear over a 50 dB dynamic range (see Fig. 3.7), and then fed into a bus which serviced Telephonics TDH-49 headphones. The frequency response of the matched filter/amplifier/earphone combination to a swept sinusoidal voltage of 80 dB re 0.0002 dynes/cm² is shown in Fig. 3.8.

The listening level was set at 80 dB SPL for a 1000 Hz sine wave with an RMS voltage equal to the RMS voltage for the steady-state vowel of the /baɛ/-/dae/ pair. The absolute intensity setting was determined with the aid of a Bruel & Kjaer artificial ear (Type 4153) calibrated with a Bruel & Kjaer Pistonphone (Type 4230). The intensity calibration was checked prior to each session, and varied less than ±0.3 dB on a day-to-day basis. For monaural presentation, crosstalk was eliminated at the headphones by disconnecting the input to the opposite earphone.

The procedure for generating the stimuli for on-line presentation was as follows: the formant transitions of the signals being combined (e.g., /b/-/d/, /r/-/l/ etc.) were loaded into core and scaled by factors of α and $1-\alpha$ which were read in from a file of randomized numbers. There were 21 stimuli, representing 21 equal steps of $\Delta\alpha = 0.05$ from

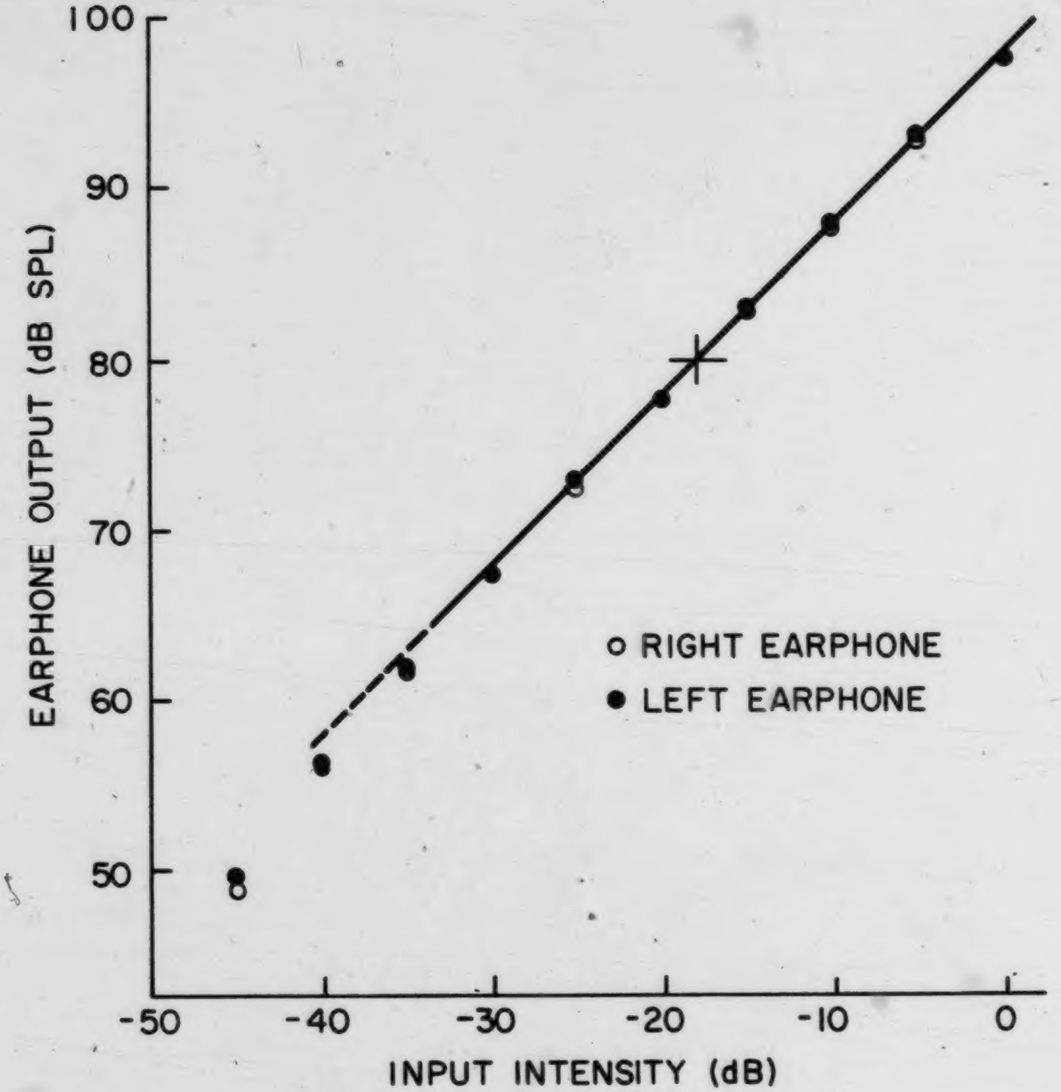


Fig. 3.7. Response of the two Braun amplifiers to a 1000 Hz sinusoid. The cross indicates the operating point for all experiments

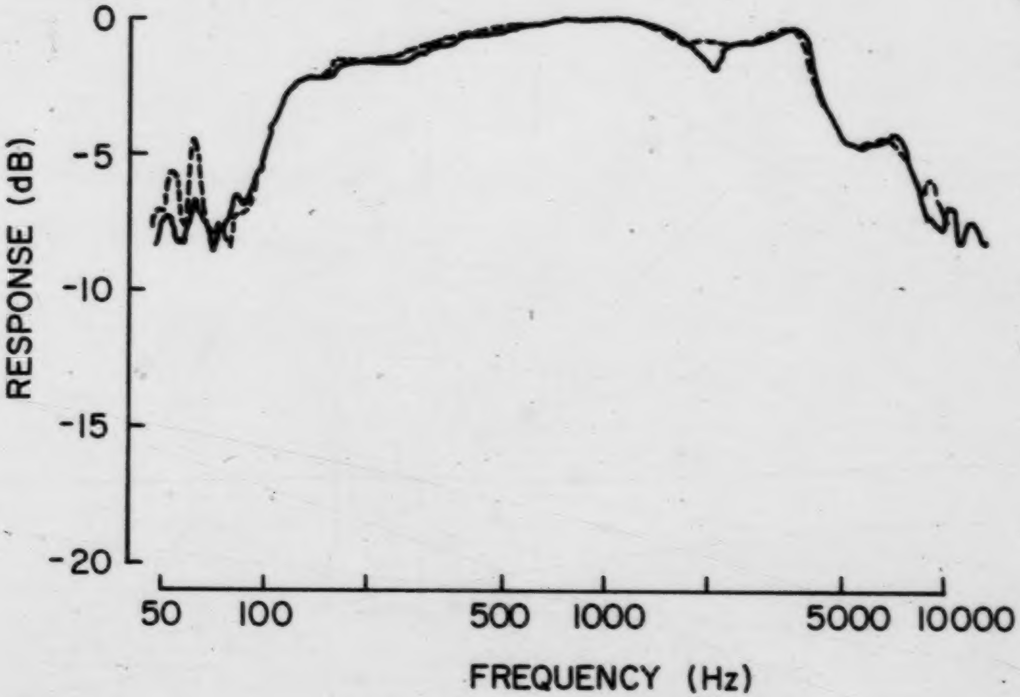


Fig. 3.8. Frequency response of the left (dashed line) and right (solid line) Rockland filter, Braun amplifier and TDH-49 headphone combinations

0 to 1. The scaled formant transitions were then added together point-by-point according to Equation 3-3 above. The steady-state vowel was then loaded into core from the disk file, and concatenated to the composite formant transitions, and the entire signal was then played back. The time required for the loading, scaling, addition and concatenation was approximately 1.5 seconds. After the stimulus was played back, switches at the remote listening stations were monitored for the subjects' responses. The program waited until all subjects had pressed one of their switches before proceeding with the next presentation. The interstimulus interval was thus somewhat variable, but averaged around four seconds. The identity of the stimulus presented and the subjects' switch choices were recorded in a disk file for later processing. After a run, the number of /b/ responses were automatically tabulated, and the identification curve was displayed on a storage oscilloscope.

The stimuli were blocked into groups of 25 in order to break up the randomization which cycled every 21 stimuli (i.e., all 21 stimuli were played back before any stimulus was repeated). A one-second 1000 Hz tone was played back at the end of each block of 25 stimuli, followed by five seconds of silence. (This was necessary to distinguish between an interblock silence and a subject's switch failing to record). Each value of α was presented 10 times, for a total of 21x10 = 210 stimuli.[4]

3.2.2 EXPERIMENT 1: Identification Scores

The first set of experiments involved identifications of the /bae/-/dae/ stimulus pair. Five subjects participated in this and other studies which involved periodic testing over a period of approximately eight months. The subjects were faculty members and graduate students of the Department of Linguistics. The author participated as a subject in all tests. None of the subjects reported any hearing deficiencies.[5]

Preliminary identification runs indicated that each subject perceived each composite stimulus as either /bae/ or /dae/ with no phonetic intrusions. The α continuum[6] was strongly categorized[7] into a region $0 < \alpha < \alpha_{50}$ corresponding to /dae/, and a region $\alpha_{50} < \alpha < 1$ corresponding to /bae/, where α_{50} was the value of α at which 50 percent recognition occurred. The value of α_{50} was different for each subject and, although the stimuli had been equated for overall intensity, did not occur at $\alpha = 0.5$ for any subject. (Possible reasons for the subject difference are discussed below). All subjects found the test to be trivial. The endpoint stimuli, being naturally-spoken tokens of /bae/ and /dae/, always resulted in 100 percent identification of these stimuli. Only in five percent or fewer of the stimulus presentations did subjects experience any difficulty in deciding. Each stimulus was generally perceived as a clear instance of a /bae/ or /dae/, and little interference from the other component was evident. (Stimuli in the transition region showed a slight increase in noisiness, but even so were easily classified as either /bae/ or /dae/).

---

[4] For purposes of comparison, the same 21 stimuli were used in all /bae/-/dae/ experiments described in this thesis.

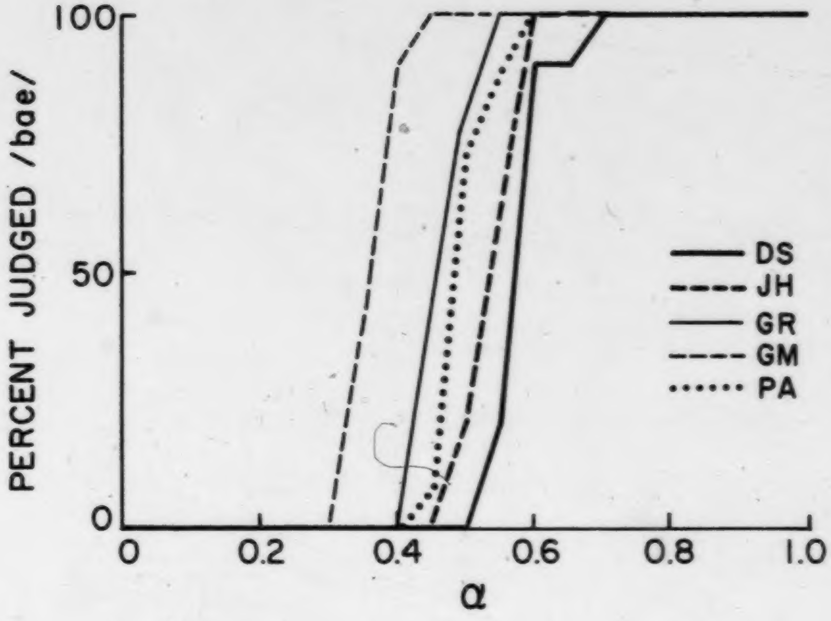
[5] Audiometric records (Appendix A) show that some of the subjects had substantial hearing losses at frequencies greater than 4000 Hz.

[6] Changing α from 0 to 1 causes a change in phonetic percept, and thus it is meaningful to speak of α as defining a "continuum".

[7] It will be shown later in this chapter that this continuum is categorically perceived.

To test the stability of the subjects boundaries (α_{50}) and also to test for ear differences, the following experiment was conducted. Each of five subjects was tested a total of three times for monaural left and monaural right conditions. The earphones were calibrated as described in Section 3.3.1 above so that the left and right earphones were as nearly matched as possible (see Fig. 3.8).

Typical identification runs for the five subjects are shown in Fig. 3.9. Fig. 3.9a shows the relative frequency of /bae/ judgements as a function of α , and Fig. 3.9b shows



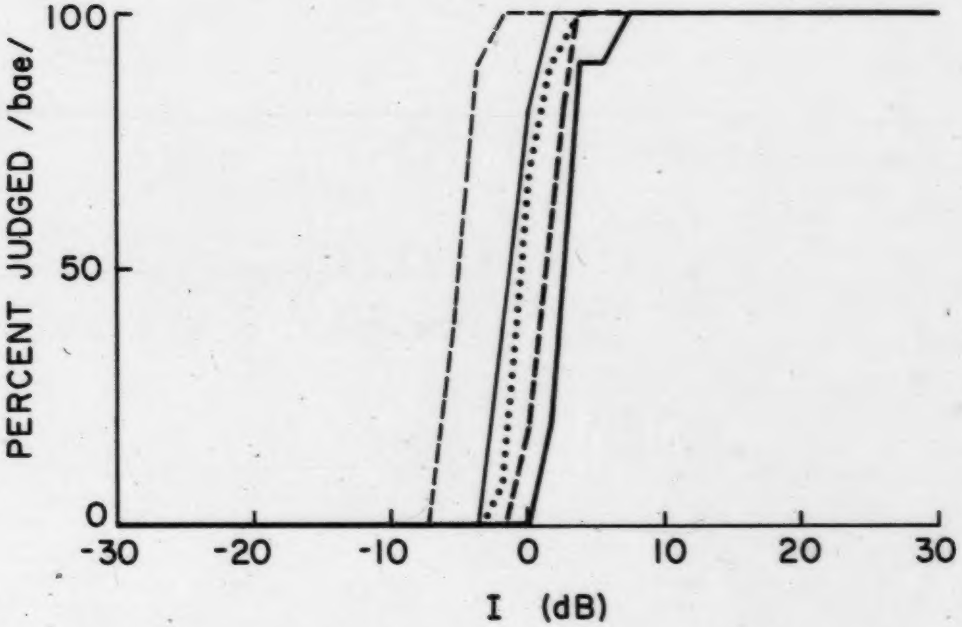


Fig. 3.9. (a) /bae/-/dae/ identification curves as a function of α. (b) The same identification curves as a function of I_α, the relative intensity of the /bae/ to /dae/ component expressed in decibels

the same data plotted as a function of I_α, the intensity of the /b/ component relative to the /d/ component, in decibels. I_α is defined by

$$I_\alpha = 20\log_{10}\left(\frac{\alpha}{1-\alpha}\right) \qquad (3\text{-}8)$$

The identification data for the left and right ear were then fitted by a least squares technique to the normal ogive

$$p(I\ ) = \Phi\left(\frac{I_\alpha - I_{50}}{\sigma}\right) \qquad (3\text{-}9)$$

The fitting process characterized each identification curve by I_{50}, the value for which 50 percent recognition occured, and σ, the standard deviation. The average boundaries for the left and right conditions are shown below in Table 3-1. A two-way ANOVA (SUBJECT x EAR) was carried out on both the boundaries (I_{50}) and widths of the transition regions (σ). Factor SUBJECT was significant for the boundaries ($p<0.001$), and accounted for 95 percent of the variance. EAR was not significant for the boundaries. Neither SUBJECT nor EAR were significant for the widths of the transition regions. There were no significant SUBJECT by EAR interactions.

TABLE 3-1

/bae/-/dae/ CATEGORY BOUNDARIES

SUBJECT	RIGHT EAR				LEFT EAR			
	I_{50} (dB)	S.D. (dB)	σ (dB)	S.D. (dB)	I_{50} (dB)	S.D. (dB)	σ (dB)	S.D. (dB)
GR	-0.69	0.52	1.18	0.17	-0.47	0.56	0.88	0.37
JH	1.18	0.61	1.26	0.24	2.02	0.19	1.18	0.56
DS	1.92	0.87	1.36	0.62	0.91	1.01	1.18	0.53
GM	-5.00	1.01	1.69	0.99	-4.72	0.65	1.07	0.30
PA	-2.28	2.74	1.75	0.69	-1.44	1.14	1.56	0.28

Identification runs were also collected over a period of several months as a result of the selective adaptation and binaural studies described later in Chapter 6. In particular, subjects DS and JH performed over 40 such identification tests, and plots of the values of I_{50} for these subjects over the testing period are shown in Fig. 3.10. It can be seen that the two subjects' boundaries changed slowly over this time period. The reason for this change is not clear, but most of these data were collected

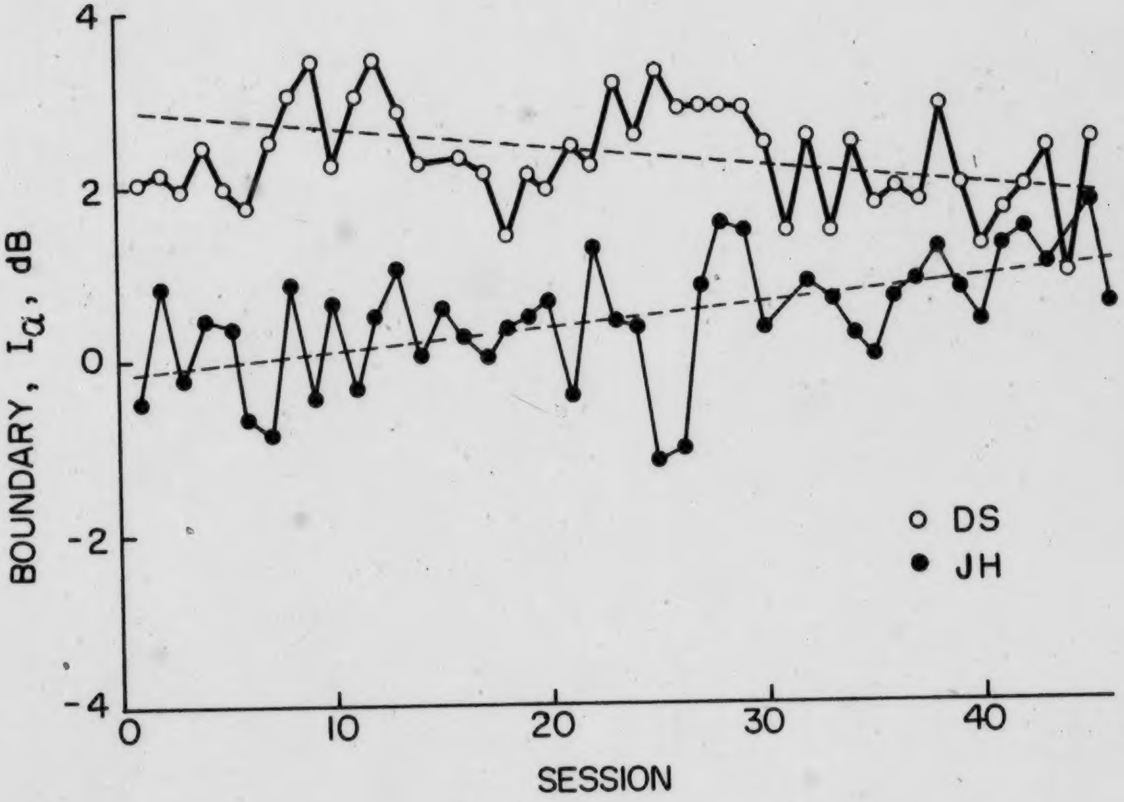


Fig. 3.10. Long-term stability of the category boundary for two subjects

during preliminary identification runs to the selective adaptation experiments described in Chapter 4. Similar stability was observed for the other subjects, who also contributed identification data spaced over several months.

The fact that the subjects show session-to-session and run-to-run variability of I_{50} which is less than the differences separating the subjects (see Table 3-1) suggests that both the boundary location (i.e., I_{50}) as well as some of the fluctuation in the boundary may be physiologically determined. The trends observed in Fig. 3.10 also suggest this. On the other hand, since some variability is observed, response bias is the most likely cause of the session-to-session variations. To obtain an estimate of the possible influence of response bias, an informal experiment was conducted using two experienced subjects. The identification test as described above was carried out twice: the first time, the subjects were required to respond "d" to a given stimulus either if it was a clear /d/, or if they thought it was a boundary /b/. On the second run, the reverse task was demanded: they were only to respond /d/ if the stimulus was a non-boundary /d/. The results for these two subjects are shown in Fig. 3.11. The amount of boundary shift is quite large (3 dB for subject DS and 6.5 dB for subject JH), and it appears that a change in response bias may be sufficient to account for the large subject differences observed in Fig. 3.9. However, it must be

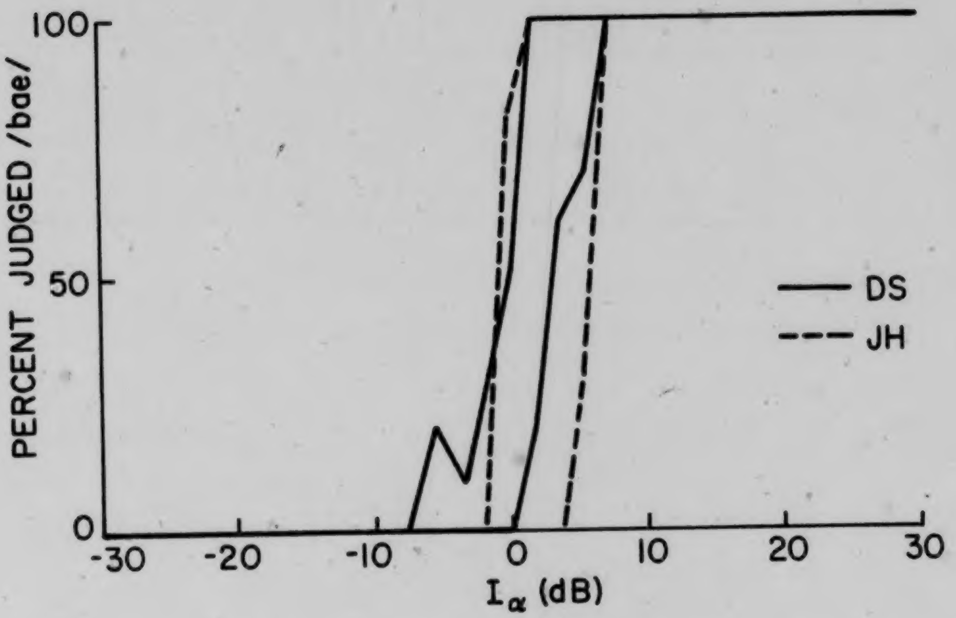


Fig. 3.11. Extreme limits of the /bae/-/dae/ identification curves for two subjects

pointed out that this experiment did not constitute an ordinary identification test. Some stimuli were called /b/ when it was perfectly obvious that under normal circumstances it would have been called /d/ (and vice versa). Thus, the changes in boundary shown in Fig. 3.10 are much greater than the variability observed between sessions (see Fig. 3.9 and Table 3-1). It is undoubtely true, however, that some of the session-to-session variability is caused by a change in response bias. However, some of the variability must be based on physiological mechanisms over which the subject has no overt control.

Hearing differences between subjects may also account for intersubject differences. Audiograms (Appendix A) show that two of the subjects (DS and PA) have regions of low spectral sensitivity, but these dips are for the left ear (most of the experiments were conducted using right ears only), and only at frequencies greater than 4000 Hz. Since the audiograms only show sensitivities at seven frequencies, no conclusive statements can be made regarding the effects of reduced spectral sensitivity in the vicinity of F_2, but since some of the observed differences are of the order of 5 dB or more, it is possible that hearing differences could account for some of the differences in the subjects' boundaries. But, since no significant difference was found between ears, and the audiograms show ear differences as pronounced as the subject differences, this contribution may be minimal.

3.2.3 EXPERIMENT 2: Identification of /bεt/-/dεt/

To test the generalizabilty of the /b/-/d/ categorization with regard to following vowel[8], tokens of /bεt/ and /dεt/ were recorded by subject JH. (The change of speaker was to test the robustness of the effect with regard to articulatory idiosyncracies). The recording procedure and stimulus preparation was carried out for these stimuli as described above in Section 3.2 for the /bae/-/dae/ pair. The presentation paradigm was in all respects identical to that described in Experiment 1 above. The same five subjects participated, and each subject was run only once (right monaural).

Fig. 3.12 shows the results of identification runs for five subjects using the stimulus pair /bεt/-/dεt/. It can be seen that the slope of the transition region is similar to that obtained for the /bae/-/dae/ pair (Fig. 3.9 above). Curiously, it appears that the relative positions of I_{50} are approximately the same for the five subjects as for the /bae/-/dae/ experiment although the spread of boundaries is

---

[8] The effect was originally discovered using a pair of /bi/-/di/ stimuli.

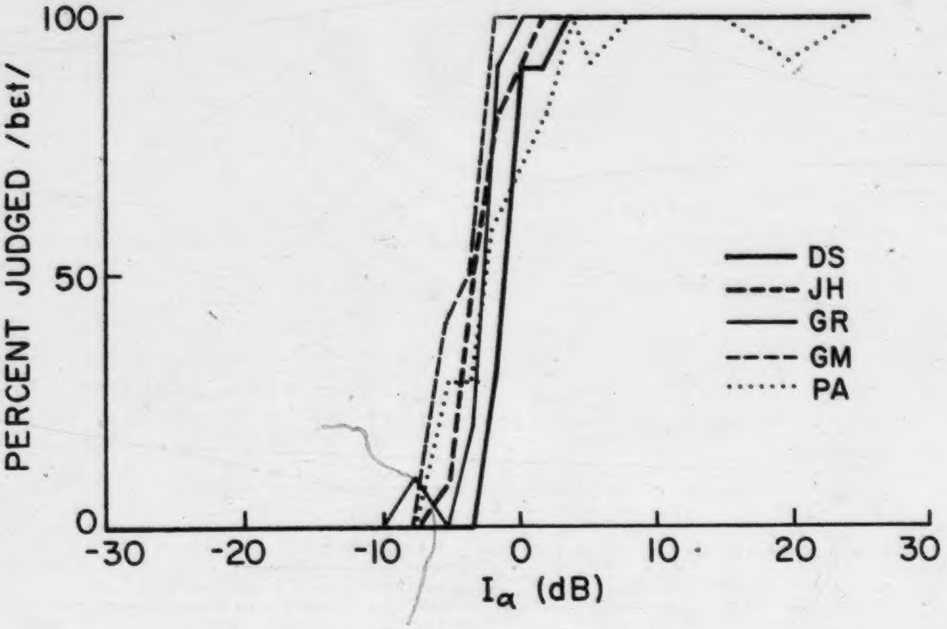


Fig. 3.12. Identification curves for /bɛt/-/dɛt/

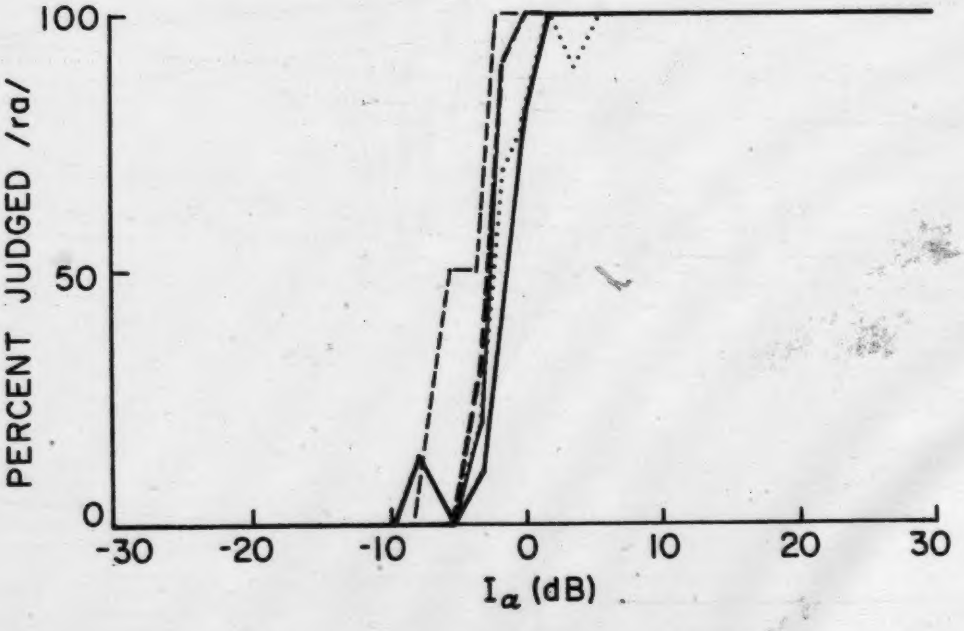


Fig. 3.13. Identification curves for /ra/-/la/

considerably reduced.[9] This finding is interesting, since it suggests that differences in hearing sensitivity may be responsible for the boundary placement. (The formant transitions for /bɛt/ and /dɛt/ are ordered similarly to those for /bae/ and /dae/).

Originally, it was hypothesized that the subject differences could be due to differential sensitivities to the various acoustic cues which are responsible for /b/ and /d/ recognition. It was felt that tokens of /b/ and /d/ from a different speaker and/or with a different following vowel might affect subjects differently. Inasmuch as only a single identification run was obtained from each subject, the present results indicate that this is not the case.

3.2.4 EXPERIMENT 3: Identification of Liquids and Vowels

To test whether a categorized continuum could be obtained with combinations formed from other than stop consonants, a /ra/-/la/ pair was constructed from a set of /ra/-/la/ tokens recorded by subject JH. The identification experiment (one run per subject) was carried out using the same presentation program and subjects as for Experiments 1 and 2. The results are shown in Fig. 3.13. Although the transition regions appear as steep as in the stop consonant test (indicating a strong categorization), the subjective impression in this case was slightly different. Whereas in the /b/-/d/ experiments little mutual interference of the two sounds was evident, in the /r/-/l/ case there was somewhat of a tendency to perceive, say, an /r/ with a low intensity /l/ in the background. Two subjects reported that additional "fusions" were heard: subject GM reported hearing both /bra/ and /bla/ and subject GR reported hearing /bla/. No such fusions were reported by the other subjects. In summary, the results of the /ra/-/la/ experiment appear to parallel the findings of /ra/-/la/ categorization by varying the F_2 -F_3 composition of synthetic stimuli (Miyawaki et al., 1975).

---

[9] Less mutual interference was noticeable for the /bɛt/-/dɛt/ stimuli than for the /bae/-/dae/ stimuli, which may be the reason for the smaller intersubject variation in I_{50}.

Considering that vowels are perceived less categorically than stop consonants (Fry et al., 1962; but note Repp et al., 1978; Pisoni, 1973; Fujisaki and Kawashima, 1969), it was decided to use the same experimental paradigm to test whether or not a vowel continuum could be created. An informal test showed that a continuum could indeed be created, but that perception was continuous rather than categorical. Two vowels /i/ and /ae/ with approximately equal f_o's were combined according to

$$V' = \alpha /i/ + (1-\alpha) /ae/ \qquad (3\text{-}10)$$

Both vowels could always be heard simultaneously. For values of $\alpha > 0.5$, the dominant percept was /i/ with a simultaneous but weaker /ae/ percept. The converse occurred for low values of α. The only stimulus combination for which the physically less intense vowel could not be heard was when $\alpha=0$ or $\alpha=1$. Since the step size was $\alpha=0.05$, this meant that the weaker vocalic percept was audible down to at least 25 dB below the stronger vocalic percept. This clearly demonstrated that little or no categorization was occurring, and this line of investigation was not pursued further. However, mixing of vowels in this fashion is a topic worthy of investigation for its own sake.

3.2.5 EXPERIMENT 4: Identification of /ba/-/da/-/ga/

A preliminary experiment showed that /ba/-/da/-/ga/ combinations produced a triply ambiguous signal. When a stimulus with approximately equal proportions of /ba/, /da/ and /ga/ was played back repeatedly, it was possible to perceive any of the three phonetic categories. Thus, in spite of the degree of spectral confusion which had to be occurring, it was clear that a three-way analogue of the /ba/-/da/ experiment was possible. If the /ba/, /da/ and /ga/ components are orthogonal, then a signal mixture given by

$$s' = \alpha s_b + \beta s_d + \gamma s_g \qquad (3\text{-}11)$$

(where $\alpha + \beta + \gamma = 1$) would result in a three-dimensional signal space as shown in Fig. 3.14. This plane should be divided into three regions, with a triply ambiguous point at the intersection of the three phonetic boundaries. The degree of orthogonality of the three-signal mixture will be reflected by the amount of interference along the phonetic boundaries.

A sequence of /ba/, /da/ and /ga/ stimuli were recorded by subject JH and the formant transitions for the /b/, /d/ and /g/ were separately stored in a disk file along with the extracted steady-state vowel from the /ba/. The /da/ and /ga/ were aligned to the /ba/ as previously described in Section 3.1. A randomized file consisting of a triangular design ($\alpha_i , \beta_i , \gamma_i$) was created, where $\alpha_i + \beta_i + \gamma_i = 1$ The scaling factors α_i, β_i and γ_i were varied in steps of 0.05. This resulted in 231 combinations of $0 < \alpha_i , \beta_i , \gamma_i < 1$. The stimuli were presented as in Experiment 1, except that a three-way response was required. Three push-buttons were provided, labelled /ba/, /da/ and /ga/. The stimuli were presented in blocks of 25, and on each presentation the subject was required to press the appropriate switch. During each run, each stimulus combination ($\alpha_i , \beta_i , \gamma_i$) occurred only once. Two subjects, JH and DS were run each a total of 10 times, resulting in 10 judgments for each of the 231 stimuli.

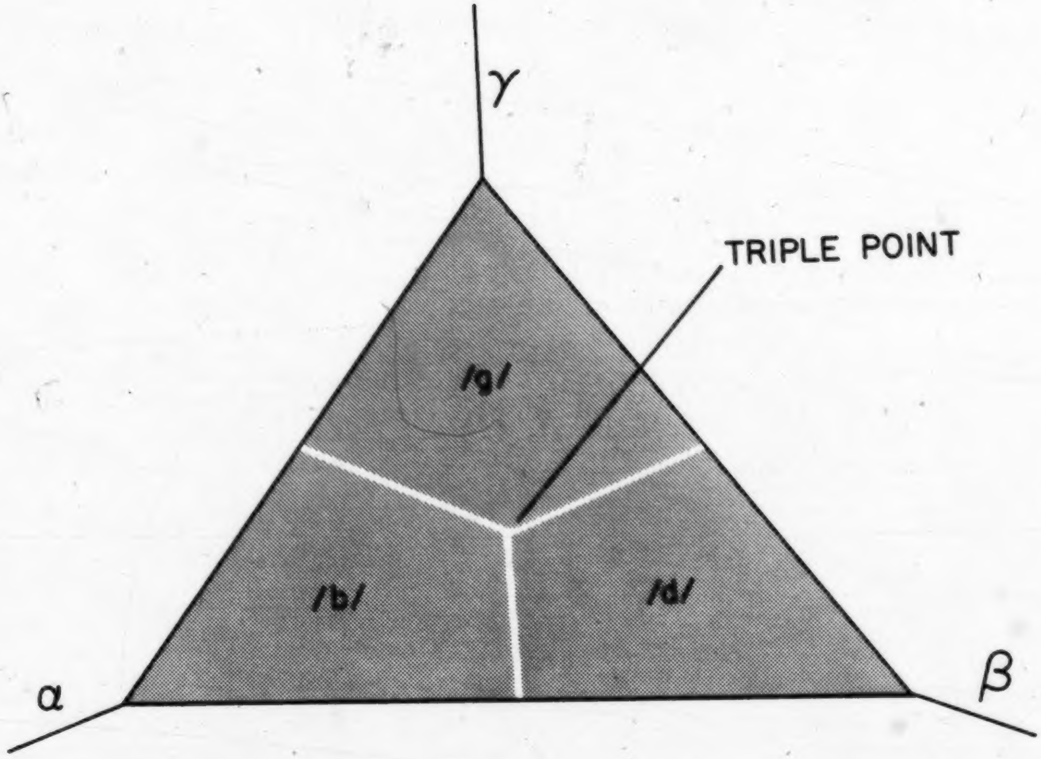


Fig. 3.14. Division of hypothetical signal space into regions of /b/, /d/ and /g/ identification

The results for the two subjects are shown in Fig. 3.15. The three symbols represent the modal values of the 10 judgements at each point. The clear division into three distinct regions with only minor irregularities around the boundaries between categories (and the fact that the boundaries lie along a line passing through the triple point and the opposite vertex) indicates that the /b/, /d/ and /g/ components of the stimuli were substantially orthogonal. Only the /g/-/b/ boundary shows a disturbance, as might be expected on the basis of spectral confusions (Cutting, 1976). By and large, the stimuli are perceived as clear exemplars of either /ba/, /da/ or /ga/, except in the vicinity of the "triple-point" where considerable noisiness was evident (more than was observed in Experiment 1).[10]

Fig. 3.16 shows identification runs which correspond to slices through Fig. 3.15 for subject DS for constant values of γ (i.e., increasing /g/ component). It is seen that for $\gamma < 0.33$ (approximately), little influence of the /g/ component is observed on the /b/-/d/ boundary. These /ba/-/da/ identification curves show transition regions with a

---

[10] This experiment was originally attempted using a /bae/-/dae/-/gae/ triple similar to the /bae/-/dae/ pair used in Experiment 1. However, after several runs had been conducted it became apparent that the /gae/ could be easily identified by its slight glide, /g^{y}ae/. To eliminate this glide - which appeared to be an artefact of articulation which additional recording attempts did not eliminate - subject JH recorded tokens of /ba/-/da/-/ga/, for which the glide was less noticeable.

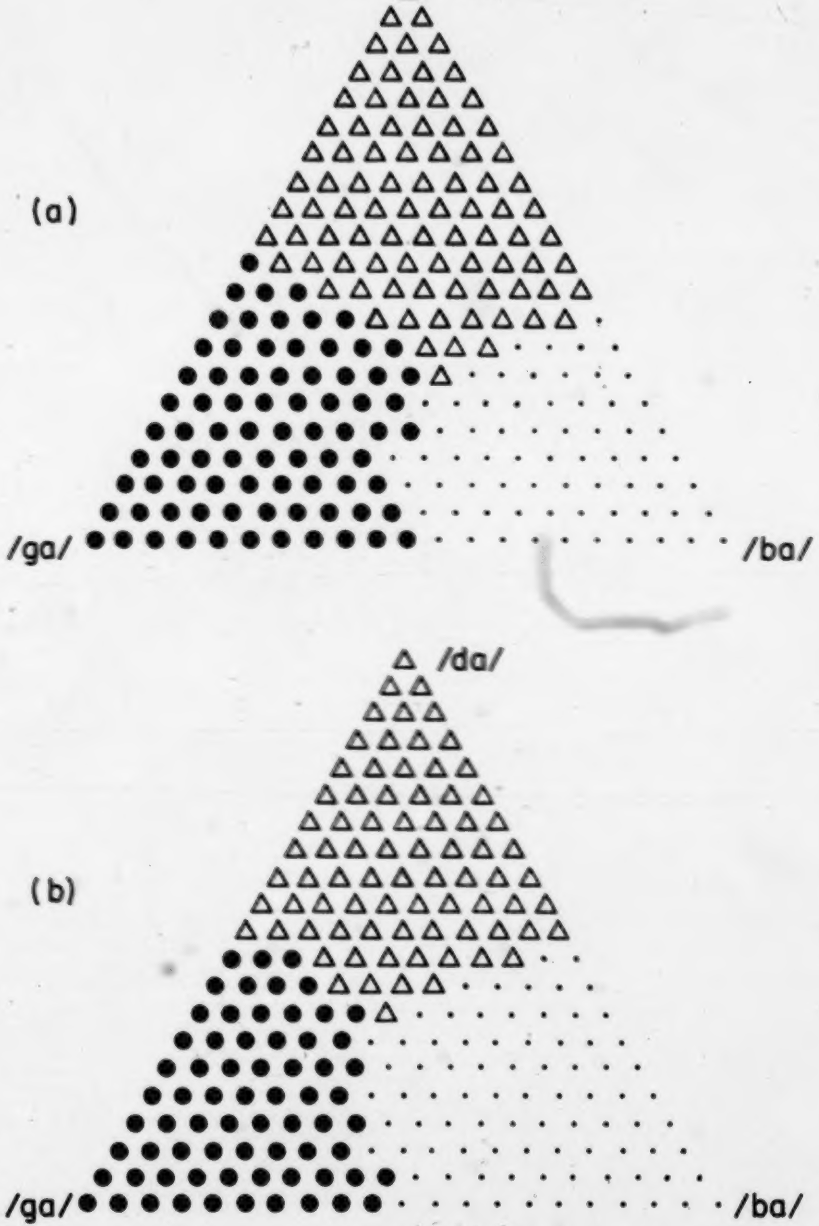


Fig. 3.15. Results of the /ba/-/da/-/ga/ identification tests for two subjects. (a) subject JH (b) subject DS

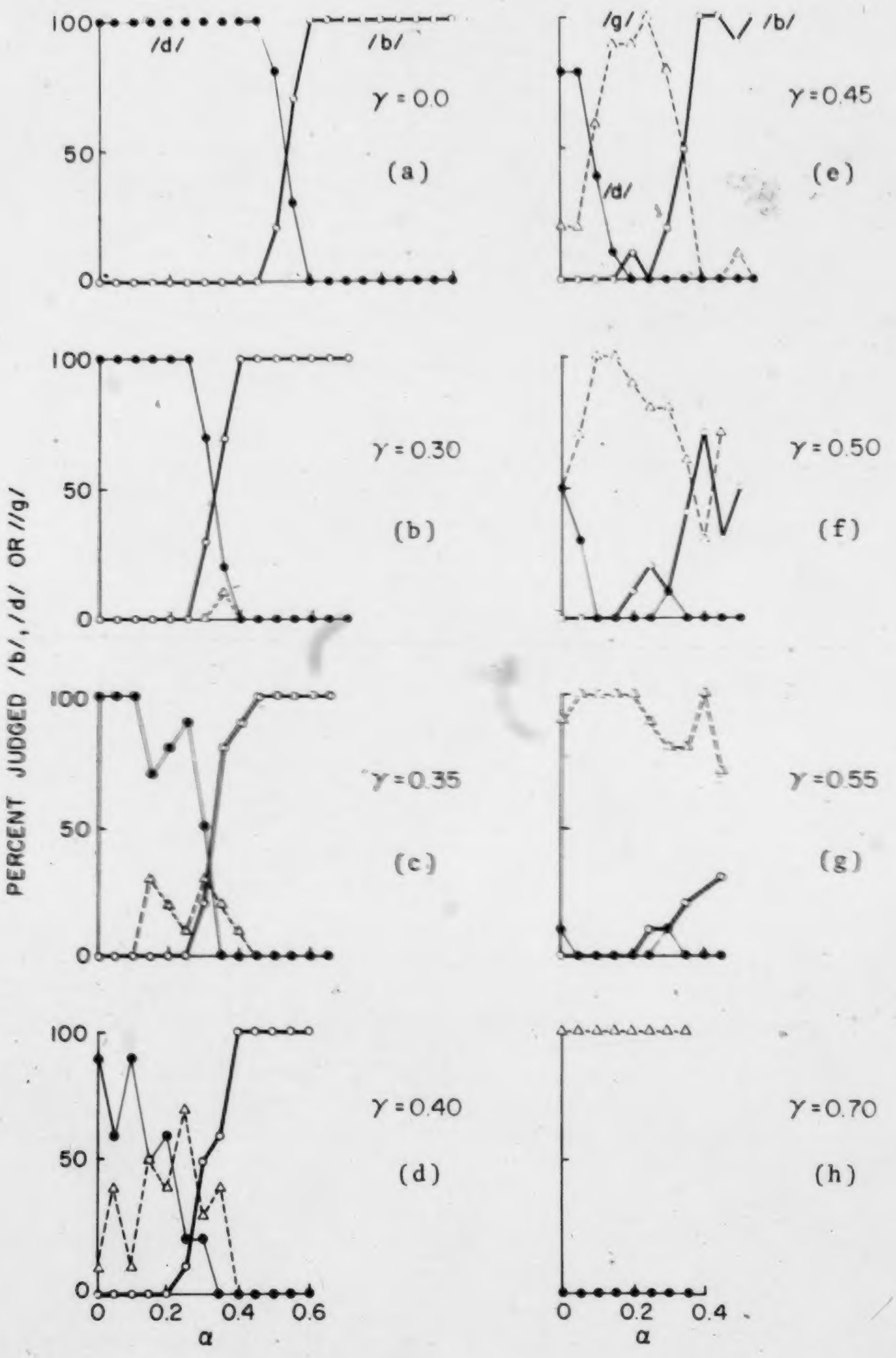


Fig. 3.16. /ba/, /da/ and /ga/ identification curves as a function of increasing /ga/ component (γ)

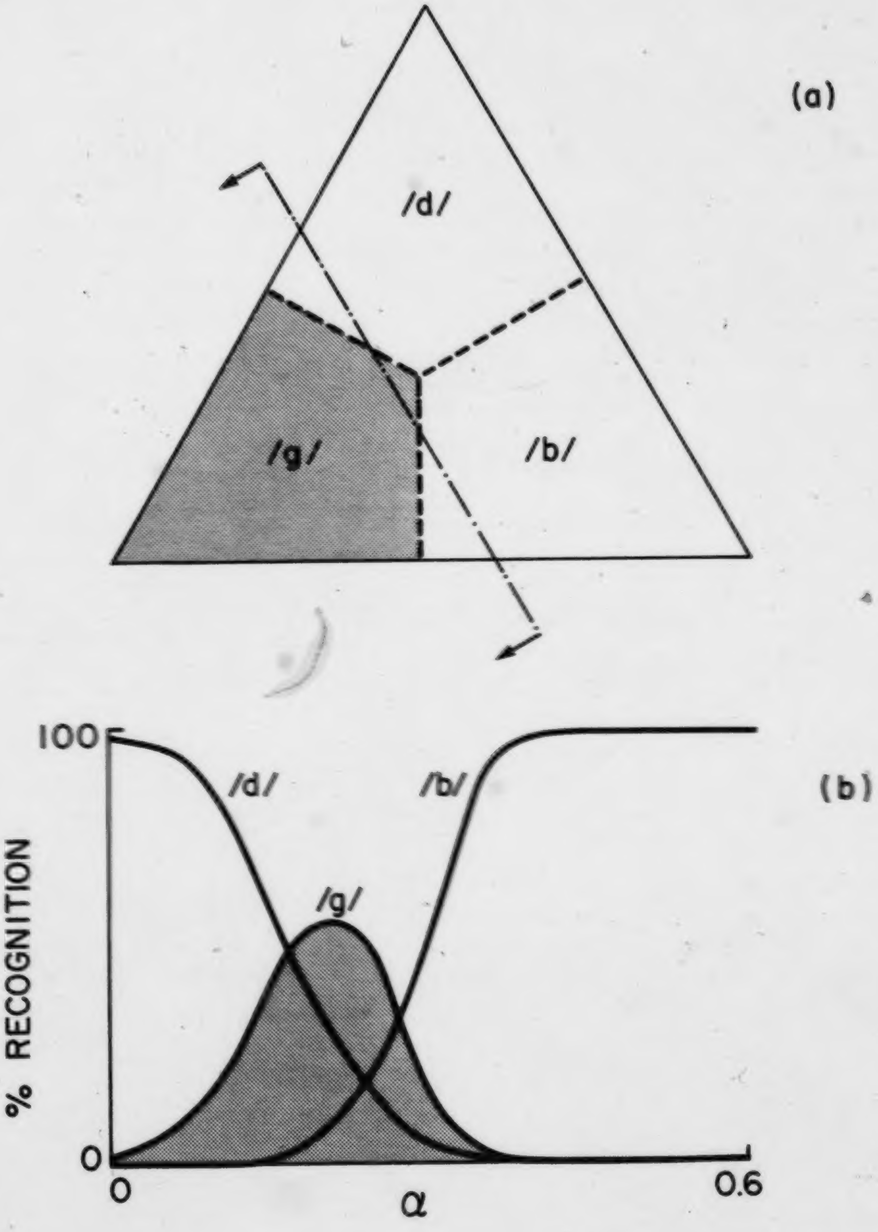


Fig. 3.17. Graphical interpretation of the results shown in Fig. 3.16. Compare (a) with Fig. 3.15 and (b) with Fig. 3.16(d)

slope approximately the same as previously observed in Experiments 1 and 2, so the addition of the /g/ component (with a simultaneous reduction in both /b/ and /d/ intensities) produces little interference until it is approximately as intense as the /b/-/d/ mixture. Although it is difficult to tell from these curves (since none of them represent a slice parallel to the phonetic boundaries), there does appear to be an increase in the number of /d/ responses at the /b/-/g/ boundary. The curious shape of the sections of constant /g/ in the vicinity of the triple point is primarily due to the fact that two phonetic boundaries are being intersected (compare Fig. 3.17 with Fig. 3.16d).

3.2.6 Summary of Identification tests

The identification tests of experiments 1 through 4 were trivial for all subjects, and little improvement (reduction in slope of the transtion region) was observed with experience of the subjects. Several linguistically naive subjects were also tested, and yielded identification data comparable in all respects to those shown in Fig. 3.9. All identification curves were sigmoidal, and rarely anything but strictly monotonic. Deviations from 100 percent recognition of non-boundary stimuli usually could be attributed to erroneous switch pressing or distraction/day-dreaming of the subject. Boundary stimuli were in general easy to label as belonging to one category or the other, although increased noisiness was evident for these stimuli. Repeated playback of a boundary stimulus creates a curious effect. It is possible to hear either /bae/ or /dae/ from such a stimulus, as the listener so chooses. It is not possible, however, to hear both simultaneously.[11] The effect can be likened to an "auditory Necker cube" phenomenon, where either of two forms can be perceived, but never both simultaneously. The effect persists even when three stimuli, /ba/, /da/ and /ga/ are combined (Experiment 5 below). The binaural counterpart of this phenomenon has been noted; Ades (1974), investigating simultaneous dichotic adaptation points out that

> "... it is worth mentioning that when /bae/ is presented to one ear and /dae/ to the other at the same time, the subject will hear a single fused percept. The fused percept may be heard as either a /bae/ or /dae/, but it is quite impossible to hear the two inputs as separate entities." (p. 612)

This is not what would be expected if the effect depended solely on masking. The almost complete suppression of the weaker phonetic percept suggests that something more is involved. If simultaneous masking were the dominant psychophysical process, it is conceivable that the influence of the masker should grow steadily with its intensity, rather than abruptly as was found in Experiments 1, 2 and 4. In actual fact, interference is only noticeable when one of the components is within a few dB of the other, i.e., in the transition region between the categories. Even then, the effect is one of added noise, not one of simultaneity of percepts. To further investigate the importance of masking in the categorization of the α continuum, an experiment was conducted using a masker which was acoustically similar to both /b/ and /d/, but which was phonetically distinct.

---

[11] Two of the subjects with extensive phonetic training noted that stimuli in the transition region tended to be slightly asynchronous (/ dae/). This slight asynchrony was not noted for the /bɛt/-/dɛt/ stimuli.

3.2.7 EXPERIMENT 5: Masking by a Vocalic Masker

A masking stimulus was created by replicating one of the pitch periods of the steady-state vowel /ae/ in order to produce a signal which provided a similar spectro-temporal structure in the vicinity of the formant transitions of the /b/ and /d/. The amplitude envelope of the /ae/ masker was modified so that the intensity ($\sum s_i^2$) of each of the pitch periods of the masker was identical to the corresponding pitch period of the /b/. This signal was aligned to the /b/ transitions by the procedure described in Section 3.2. The resulting signal, when concatenated to the steady-state vowel from which it was obtained, produced a natural sounding /ae/. This masker was combined with the /b/ and /d/ formant transitions to create two series of stimuli:

$$b' = \alpha/b/ + (1-\alpha)/ae/ \qquad (3\text{-}12a)$$

and

$$d' = \alpha/d/ + (1-\alpha)/ae/ \qquad (3\text{-}12b)$$

Forty-two stimuli were used, twenty-one each of b' and d'. Thus, $\alpha = 0$ corresponded to a /ae/ for both sets of stimuli. These were included to obtain a measure of the response bias in favour of either /b/ or /d/. The two series of stimuli were presented in random order in a fully-crossed design. All aspects of the stimulus presentation were identical to that of Experiment 1. The subjects' task was to identify each stimulus as /bae/ or /dae/. The subjects were fully informed as to the nature of the experiment and were asked to try as hard as they could to identify whichever stimlus (/b/ or /d/) was being presented. It was specifically pointed out that the /b/ and /d/ stimuli occurred equally often.

The number of /b/ and /d/ responses as a function of were calculated and plotted in Fig. 3.18. (The I_α scale for each subject was adjusted by I_{50} dB, i.e.,

$$I'_\alpha = I_\alpha - I_{50} \qquad (3\text{-}13)$$

where I_{50} is the average boundary as determined from

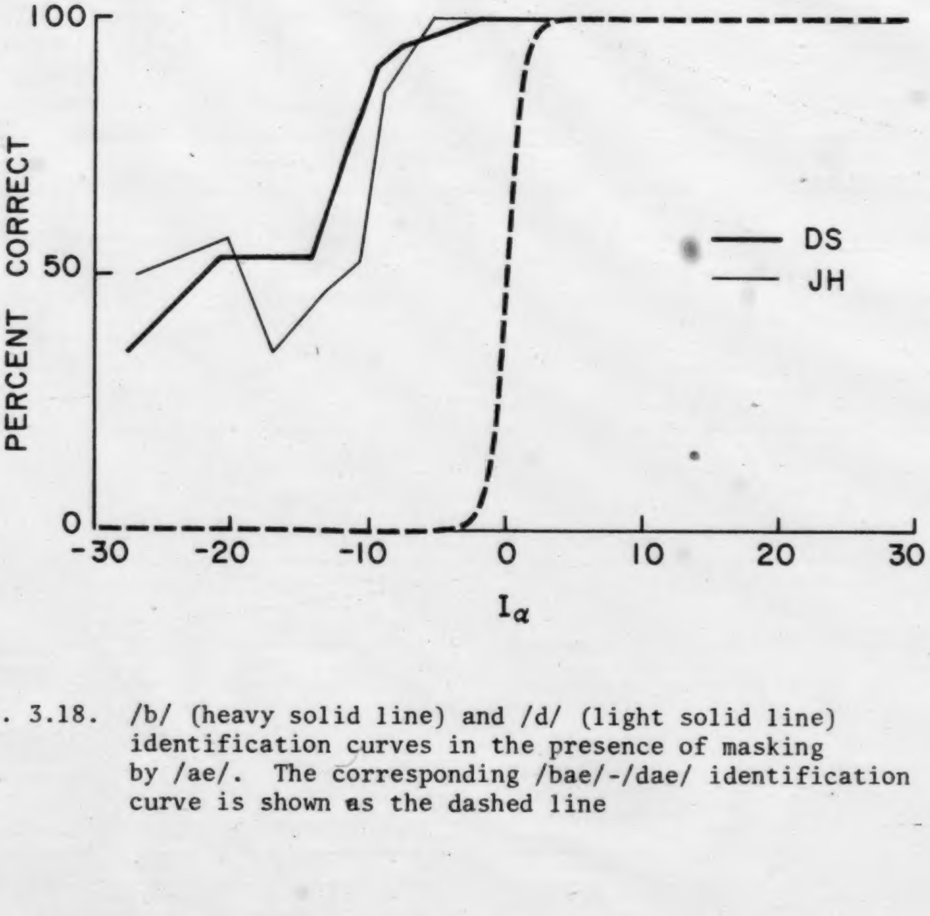


Fig. 3.18. /b/ (heavy solid line) and /d/ (light solid line) identification curves in the presence of masking by /ae/. The corresponding /bae/-/dae/ identification curve is shown as the dashed line

Experiment 1). For purposes of comparison, the average /b/-/d/ identification curve from Experiment 1 is also shown. The results show that the /ae/ "transitions" masked the /b/ and /d/ transitions to an extent which monotonically increased with the strength of the /ae/ masker (compare with Fig. 3.9). Furthermore, 100 percent identification of both /b/ and /d/ occurred 10 dB or so before the /bae/-/dae/ boundaries obtained in Experiment 1, which indicates that the mutual masking of the /b/ and /d/ is more effective than that of the /ae/ masker. This is not what would be expected if spectral masking were the only factor involved.

One possibility is that some level of phonetic- level inhibition is occurring, in which the outputs of some phonetic processors interact in such a way that only the strongest excitation is perceived. This is consistent with the notions of "rapid encoding" of consonantal cues (Pisoni, 1973, 1975; Fujisaki and Kawashima, 1969). It may be that the abrupt change of percept is related to the fact that the acoustic cues for /b/ and /d/ develop over the same amount of time, in which case some form of mutual interaction during the processing of these two signal components may account for the apparent inability of a subject to perceive both percepts simultaneously.[12]

3.3 ABX AND AX DISCRIMINATION TESTS

The identification results of Experiments 1 as well as informal observations indicated that the /b/-/d/ combinations were strongly categorized. To test whether or not categorical perception was occurring (according to the contemporary criteria set forth by Studdert-Kennedy et al., 1970), a conventional ABX discrimination paradigm was performed using the /bae/-/dae/ stimuli.

3.3.1 EXPERIMENT 6: ABX Discrimination

Twenty-one stimuli were prepared by computing the various combinations for α ranging from 0 to 1 in steps of 0.05, and each mixture was then scaled by $1/\sqrt{I'}$, where I' is given by

$$I' = 1.00 - 1.36\alpha + 1.36\alpha^2 \qquad (3\text{-}14)$$

---

[12] An informal experiment was conducted with a mixture of /bae/ and /rae/. In this case, a transition from /b/ to /br/ to /r/ was observed, similar to phonological fusions in dichotic studies (see Cutting, 1974, 1976). This supports the notion that the interaction between /b/ and /d/ depends in part on their acoustic similarity.

This scaled all signal combinations to the same overall intensity.[13] These 21 stimuli were stored in a disk file and accessed by the presentation program described below.

The experimental setup for the discrimination experiment was as described for Experiments 1 through 4. For each presentation, two stimuli A and B which were separated by $\Delta\alpha = 0.1$ (i.e., two steps) were selected according to a file of randomized numbers. The third stimulus, X, of the ABX paradigm, was always either A or B. The interval between A and B, and B and X, was 750 milliseconds. Subjects were required to press switch 1 if the third stimulus was judged to be the same as the first stimulus, or switch 2 if it was the same as the second stimulus. Three subjects participated in this test, with each subject repeating the discrimination run five times. From the five runs, 40 judgements of the discriminabilty of each pair of stimuli were obtained.

The results of the ABX discrimination task for the three subjects are shown in Fig. 3.19. The data for each subject have been normalized to a common category boundary using Equation 3-13.

---

[13] A preliminary run showed that the overall intensity difference of the composite stimuli as a function of the weighting parameter α (see Fig. 3.4) aided discrimination. Scaling by $1/\sqrt{I'}$ eliminates this variation with α.

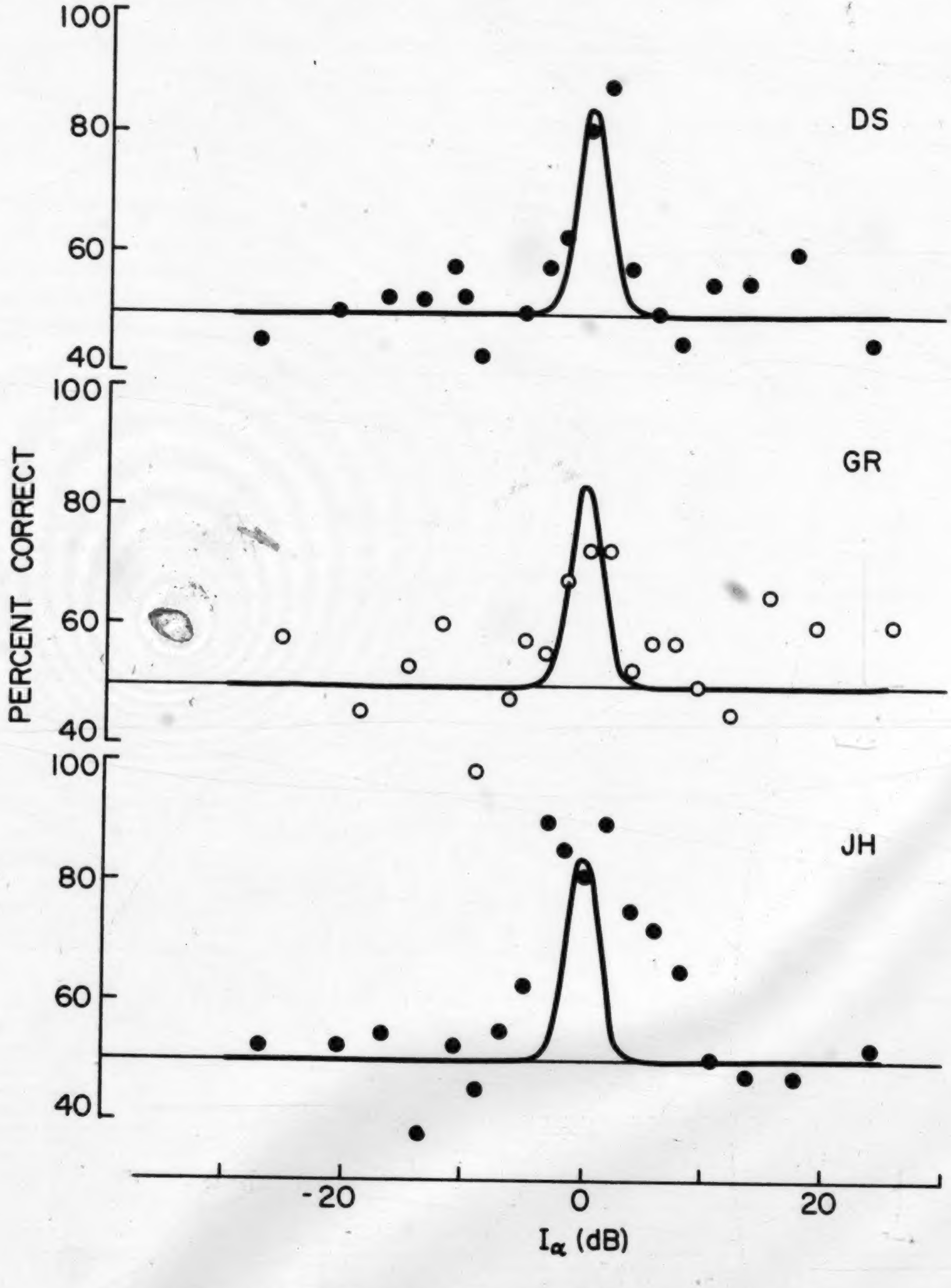


Fig. 3.19. ABX discrimination curves for the three subjects. The solid lines are calculated from Equation 3-1

Inasmuch as all three subjects had by this time become very acquainted with this set of stimuli, discrimination was a very difficult task, and stimuli could be judged "different" with relative confidence by a subject only in approximately 5 percent or fewer cases. This meant that most of the time the subject was required to attempt discriminations between stimuli for which (or so he felt) he could not perceive any difference. This made the task very difficult for the subjects, for it was difficult for them to know when their responses were at all systematic. The results show considerable statistical scatter even when five runs are averaged, but nonetheless show enhanced discriminability in the vicinity of the phonetic boundary. For all three subjects, the discrimination curves asymptote in the wings ("troughs", but "wings" is more appropriate) to chance level (50 percent).

Discrimination data are typically compared with the Haskin's model (Equation 3-1) and/or the Fujisaki and Kawashima model (Equation 3-3) to determine whether or not the continuum is categorically perceived. In the present instance, only the Haskin's model can be applied since the discrimination curves must go to chance level at the endpoints. This follows from the way in which the stimuli are constructed. Near $\alpha = 0$ or $\alpha = 1$, the stimuli become increasingly pure, and hence discrimination must go to zero. The labelling probabilities, $p(I_\alpha)$, were calculated from the normal ogives determined from Experiment 1. (p(I_α) is the probability of identifying a stimulus specified by as a /b/). The average value of σ ($\bar{\sigma}$ = 1.2 dB) was used for this calculation. The predicted functions are shown as the solid lines in Fig. 3.19. For subjects DS and GR, the enhancement of discriminability is roughly that predicted by the model. Not much can be claimed about the quality of fit, since even with five runs the data have not stabilized.[14] Nonetheless, a strong peak in the vicinity of the category boundary is evident. The data for subject JH show a very poor fit, with the measured discriminability being much greater than that predicted by the Haskin's model. This is evidently due to the fact that this particular subject chose to reduce the ABX test to an AX test (see below).

3.4 AX DISCRIMINATION

Various difficulties were encountered with the ABX paradigm described above. First of all, the task was very difficult. It was felt by all three subjects that for the most part responses were being given randomly. While this may not be important for the paradigm, it is important for the subjects concentration, since it is difficult for subjects to respond consistently when no basis for consistency can be perceived. Second, it was apparent during the ABX discrimination experiment that subtle differences in the stimuli which were noticeable during an identification test were no longer perceivable. Third, it was discovered that subjects JH and GR had effectively short-circuited the ABX paradigm. According to their confessions, they chose to monitor only to the last two (B and X) stimuli, and if they were different, they responded as if the first and third stimuli were identical. Likewise, if the last two stimuli were judged to be the same, they responded as if the second and third stimuli were the same. (Only the author, subject DS, naively attempted to compare all three stimuli). Inasmuch as the nature of the task had been explained to the subjects at the outset, they cannot be faulted for having chosen to make life easy for themselves. It just stresses one of the prime disadvantages of the ABX paradigm, i.e., that there is more than one possible subject strategy (see MacMillan et al., 1977; Pollack and Pisoni, 1971; Pierce and Gilbert, 1958, for other possible subject strategies).

---

[14] The data for these three subjects cannot be pooled since their identification boundaries are separated by several dB. The transformation $I'_\alpha = I_\alpha - I_{50}$ causes the data points to correspond to different values of I'_α .

To overcome the difficulties of the ABX paradigm, and to obtain data to test the AX discrimination model presented in Chapter 2.3, discrimination testing was continued using a fixed-standard AX paradigm similar to that used by Carney et al. (1977). The comparison between the STD model and the phonetic memory model indicates a number of possibilities for testing the relative adequacy of these two models within the AX testing paradigm. The behaviour of the STD model as a function of the observer criterion is considerably different from that of the phonetic-memory model, and consequently the change in discrimination scores with changes of observer criterion is of considerable interest in the verification of the model.

3.4.1 EXPERIMENT 7: AX Discrimination Scores

The same stimuli as in Experiment 5 were used in a fixed-standard AX paradgim. On any particular run, one of these stimuli was chosen as the standard against which all other 21 stimuli were compared. The order of the standard and test stimuli were randomized, and the interstimulus interval was set at 500 milliseconds. Four subjects were employed, each providing seven individual runs. Only a single subject was tested at one time in order to reduce the number of distractions in the environment. The subjects were fully informed of the composition of the standard stimulus and the proportion of true same-same contrasts.[15]

The first run was used as a training run to familiarize the subject with the experimental task. This introduction to the test was necessary to ensure that the subject clearly understood what was meant when he was asked to keep the same criterion both within and between runs. All subjects chose a fairly lax criterion for this first run, attending mostly to phonetic differences. On later runs, they were instructed to adopt either stricter or more lax criteria.

Each subject was run at least a total of seven times. On six of these runs, three different standards were used: $\alpha = 0$ (pure /dae/), $\alpha = 1$ (pure /bae/), and a value of close to the subjects phonetic boundary as determined by a preliminary identification run.[16] Two replications were obtained for each standard, and the subject was instructed to try to maintain the same criterion as he/she had used on the previous run. It was suggested to the subjects that in order to help maintain the same criterion, roughly the same proportion of "same" and "different" judgements should be maintained. Each standard (I_s) was replicated before a new standard was chosen.

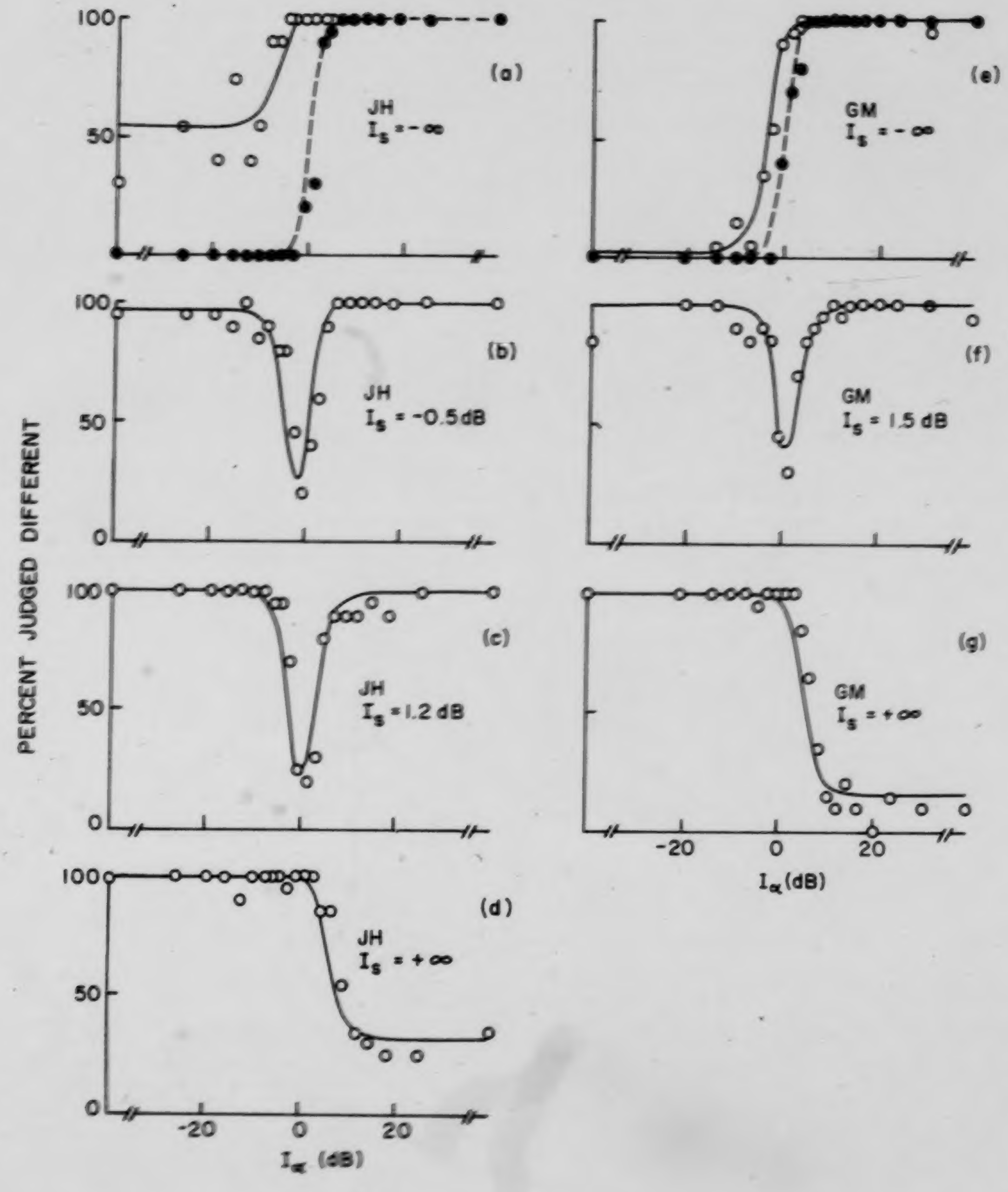
---

[15] Since the objective was to obtain information on the effect of observer criterion on the AX discrimination curves, there was no point in raising the number of AA contrasts to 50 percent, as done by Carney et al. (1977). Also, a change in observer criteria thus affects not only the number of false alarms (AA contrasts judged "different"), but also changes the number of "different" judgements across the entire continuum (see 2.11).

[16] It was not possible to use the exact boundary stimlus as a standard for all subjects since (a) this is not an exactly definable value of α and (b) only 21 stimli in steps of $\Delta\alpha = 0.05$ were available. The value of α closest to that subject's average boundary was used.

Testing was carried out on two or more days, with replications always being performed on the same day. Approximately 15 minutes to one-half hour was allowed between runs in order to minimize adaptation effects which might occur from repeated presentation of the standard (see Simon and Studdert-Kennedy, 1978). On the last run, the standard was selected as $\alpha = 0$, and the subject was asked only to respond to phonetic differences. Several additional runs were obtained from subject JH.

The results for the five subjects are shown in Fig. 3.29. The influence of observer criterion is as predicted by the AX discrimination model derived in Section 2.3. For a lax criterion (only phonetic differences), the AX discrimination curve coincides with the identification function as both the phonetic-memory and STD models predict. As the same-different criterion is tightened so that finer differences are responded to, the discrimination curve is still sigmoidal, but its slope increases and it shifts towards whichever endpoint stimulus is the standard. As the criterion is tightened further, the number of within-category responses (and false alarms) increases, also as


PERCENT JUDGED DIFFERENT
100
50
0
(a)
JH
I_s = -∞
(b)
JH
I_s = -0.5 dB
(c)
JH
I_s = 1.2 dB
(d)
JH
I_s = +∞
(e)
GM
I_s = -∞
(f)
GM
I_s = 1.5 dB
(g)
GM
I_s = +∞
-20
0
20
I_α (dB)

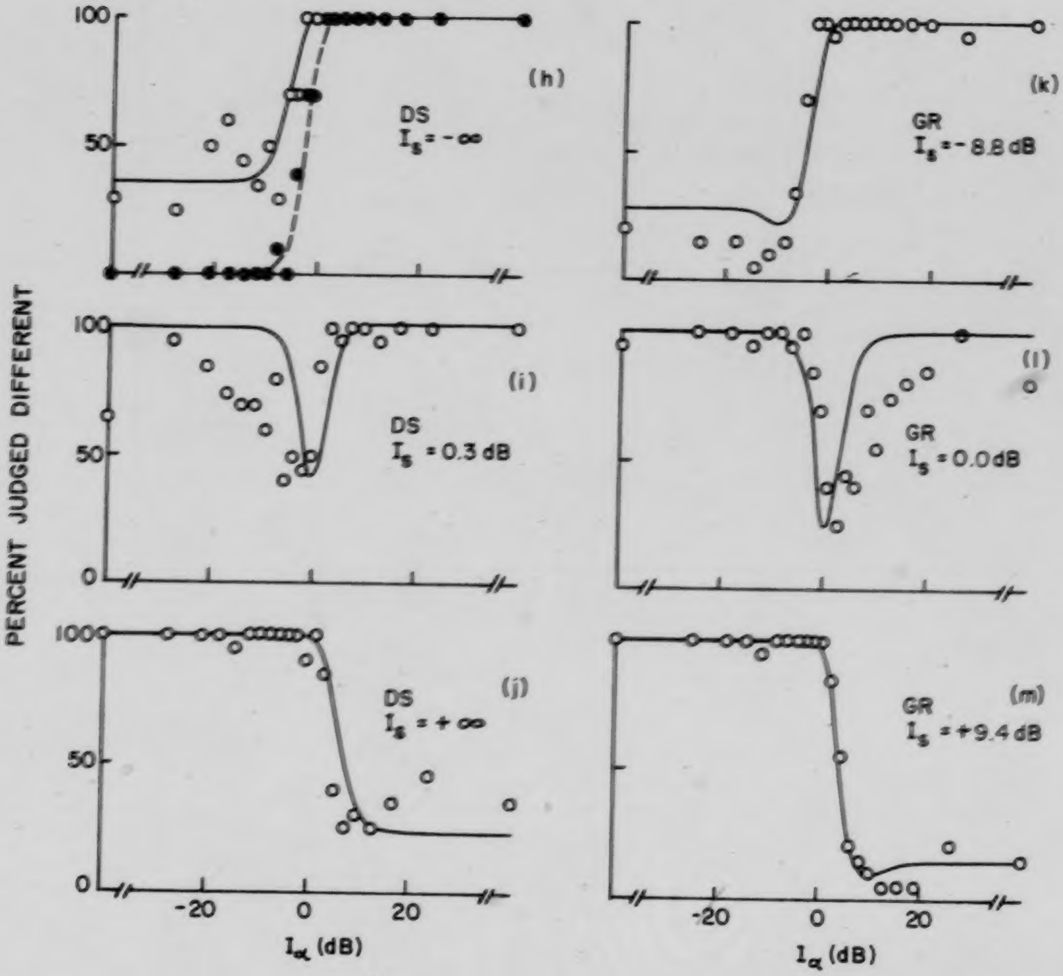


Fig. 3.20. AX discrimination results. The standard stimulus is indicated in each graph by the arrow on the horizontal axis, and also in the legend by "$I_s =$ ". The solid lines represent the fitted model

predicted. Fig. 3.21 shows the individual runs for the pure /dae/ standard stimulus plotted on a single graph. Comparing the results of Fig. 3.21 with Fig. 2.9 and 2.13, there seems no doubt that the STD AX discrimination adequately describes the influence of the observer criterion. The phonetic memory model (Equation 2-5) can be conclusively rejected since it obviously does not demonstrate this behaviour (compare Fig. 3.20 with Fig. 2.4).

When the standard stimulus in the AX discrimination task is a boundary or near-boundary stimulus, the discrimination function shows a pronounced dip at the location of the standard, as is predicted by the STD model. (For this test, all subjects were asked to respond to any differences which they might feel existed, bearing in mind that approximately 5 percent of the comparisons would be between physically identical stimuli). The symmetrical nature of these curves (Fig. 3.20b and 3.20f) indicates that a boundary stimulus is perceptually equidistant from either endpoint. This supports the notion that phonetic differences are merely "large acoustic differences" insofar as a discrimination test is concerned.

It cannot be inferred directly from these results that categorical perception is occurring, or, for that matter, that it is not occurring, since a change in observer

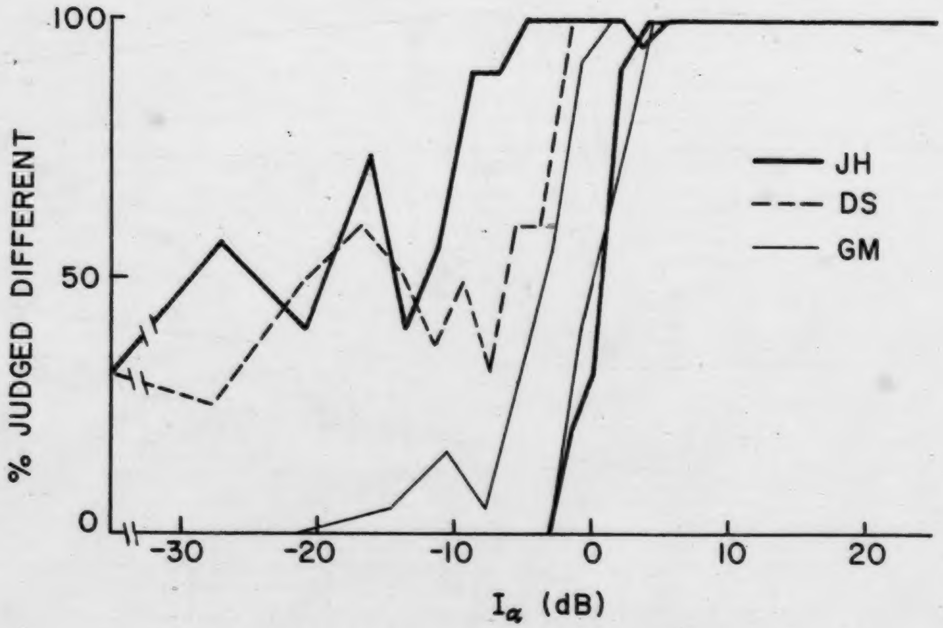


Fig. 3.21. Summary of AX discrimination curves for /dae/ standard . Compare with Fig. 2.11(b)

criterion is, to a first approximation at least, equivalent to an increase or decrease in discriminability. (Likewise, the fact that the ABX curves presented in Fig. 3.19 asymptote to 50 percent does not imply that no within-category discrimination is possible. It can equally well mean that the observer chose not to respond to slight acoustic differences). To test for categorical perception, the curve-fitting model of AX discrimination derived in Section 2.2 was fitted to the data.

3.4.2 Fitting the AX Discrimination Data

The analysis of the "dispersion model" of discrimination presented in Section 2.2 suggests that the results of Experiment 7 fit to the model if a suitable unimodal dispersion function is chosen. (For this particular continuum - relative intensity - it is not apparent why dispersion should be unimodal, but the results clearly indicate that it is). A convenient function to test out the dispersion function is a Gaussian, as described previously in Section 2.2. The fitting process was carried out by first averaging the replicated runs for each of the subjects, and then adjusting all I values by Equation 3-13 so that the identification boundary for all subjects occurred at $I_{50} = 0$ dB.[17] This reduces the Gaussian dispersion curve to a single free parameter since the mean is then positioned at $I_{50} = 0$ for all subjects. To fit the individual discrimination profiles, it was necessary to allow the criterion Δy_c to be fitted to each discrimination run independently.

---

[17] The justification for this is provided below.

The model to which the data were fitted are defined by Equations 2-11, 2-13 and 2-15:

$$P_D = 1 - \Phi\left(\frac{y_2 - y_1 + \Delta y_c}{\sqrt{2}\,\sigma}\right) + \Phi\left(\frac{y_2 - y_1 - \Delta y_c}{\sqrt{2}\,\sigma}\right) \qquad (3\text{-}15a)$$

$$y = \int_{I_{50}}^{I_\alpha} D(I_\alpha)\,dI_\alpha \qquad (3\text{-}15b)$$

$$D(I_\alpha) = \frac{1}{\sqrt{2\pi}\,\sigma_D} e^{-\frac{(I_\alpha - I_{50})^2}{2\sigma_D^2}} \qquad (3\text{-}15c)$$

The fitting process consisted of varying (a) the "width" (i.e., standard deviation) σ_D of the Gaussian dispersion function (b) the standard deviation σ of the random variable Y, and (c) the criterion Δy_c for each profile. The data were fitted by computing the model for each of the profiles and computing the sum of squares difference over the whole data set. Each of the parameters was incremented or decremented on each cycle according to whether or not the sum of squares difference was increasing or decreasing (basically following the PEST algorithm of Taylor and Creelman, 1966). The iteration was stopped when the sum of squares difference changed by less than 0.05 percent. The resulting model is shown in Fig. 3.22, and the fitted discrimination profiles are shown as the solid lines in Fig. 3.20. In general, the fit is quite good, considering the stability of the data. (More than two replications are obviously necessary to stabilize the data. Even five runs as done in the ABX experiment still showed large scatter). Some of the differences which exist are attributable to the fact that the values of I_{50} chosen for the standard stimulus were based on the average I_{50} for each of the subjects as determined from Experiment 1. I_{50} may be different by as much as 1 or 2 dB on any particular run. As Fig. 3.9 shows, the effect of a change in the location of the maximum of the dispersion function will have its greatest effect in the vicinity of a subject's boundary. Thus, the profiles in Fig. 3.20 may not be properly ordered with respect to I_s . This is undoubtedly the case for subjects DS and GR (see Fig. 3.20i and 3.20l). For subject DS the standard stimulus was a clear /daɛ/, which indicates that it is not as close to the boundary as Fig. 3.20i would imply. Subject GM, on the other hand, commented that she felt the standard stimulus must have been close to the boundary since she could not unambiguously identify it. Her boundary run (Fig. 3.20f) shows that this is indeed the case. Subject JH, for whom two runs were taken which spanned his normal boundary, shows a reversed symmetry, as would be expected (see Figs. 3 .20b and 3.20c).

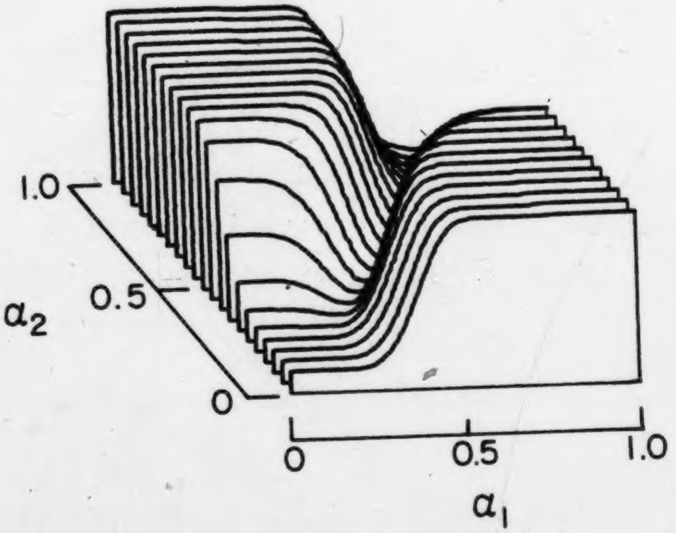


Fig. 3.22. Best fit for the AX discrimination model for $\Delta y_c = 0.2$

The normalization of the various subjects' data by shifting the discrimination curves by an amount necessary to align the boundaries at $I_{50} = 0$ dB requires justification. To check this, the fitted model was computed with the dispersion curve offset by amounts I_{50}, where I_{50} were taken from the results of Experiment 1. The discrimination profiles thus generated were compared with profiles calculated for a centered dispersion curve and shifted by the same amount I_{50}. The differences were only of the order of a few percent, which is much less than the scatter observed in the experimental data. Consequently, the data can be pooled in this fashion with only a first order loss in accuracy.

3.4.3 Summary of Discrimination Results

The STD AX discrimination model developed in Chapter 2 was fitted to the discrimination data and was found to adequately characterize the experimental data. This provides support for this model as a general model for categorical perception, and demonstrates that the relative intensity continuum under investigation is categorically perceived. The phonetic memory AX discrimination model (Equation 2-5) is evidently inadequate to account for the observed discrimination data.

Fig. 3.23a shows the dispersion function obtained from

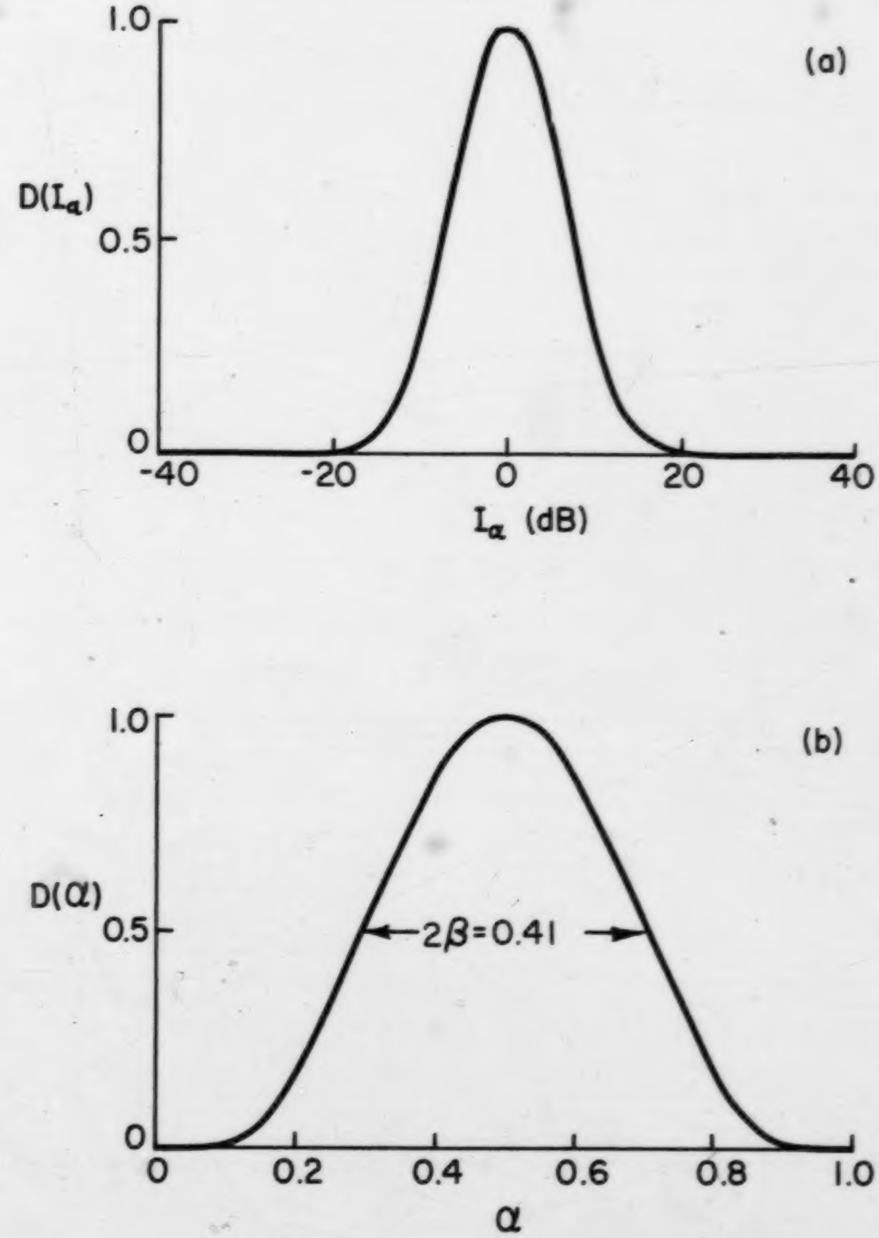


Fig. 3.23. Dispersion function for the best-fit model, plotted in (a) as a function of I_α and in (b) as a function of α. In (b), β is shown defined as the width of the dispersion function at half-height

the fitted model. Fig. 3.23b shows the dispersion as a function of the control parameter α. The measure of categoricity previously suggested (Equation 2-17) based on the α scale rather than the I_α scale (since the endpoint stimuli are an infinite distance apart on this scale) is

$$\varepsilon = \frac{1}{\beta} \simeq 2.5 \qquad (3\text{-}16)$$

Not much can be made from this index, of course, since there are no comparable indices in the literature. Nonetheless, by comparison with other categorical perception studies, this continuum appears to be as strongly categorized as typical VOT or F_2 continua, and thus values of ε greater than 10 or so would appear to typify what have been called categorically perceived continua.

3.4.4 What is Creating the Dispersion?

Inasmuch as the choice of a Gaussian function for the underlying disperion is completely arbitrary, the choice is justifiable on the grounds that it is unimodal and vanishes as $I_\alpha \rightarrow \pm\infty$. This is a necessary condition, since the endpoint stimuli for this continuum are uniquely defined by $\alpha = 0$ ($I_\alpha = -\infty$) and $\alpha = 1$ ($I_\alpha = +\infty$). The dispersion function should obviously vanish at these points. At least some of the differences between the model and the data (Fig. 3.20 above) can be attributed to the particular choice of dispersion function.

The origin of the dispersion is at present unclear. It appears to be based on the fact that the signal contains two components which are being simultaneously, plus the fact that the two component signals are shown to be categorically perceived on the F_2-F_3 continuum. The close parallel between the identification and discrimination results for /bae/ and /dae/ on this relative intensity continuum and those obtained using and F_2-F_3 continuum suggest that some common mechanism is involved. One possible explanation is that some interference, perhaps inhibitory, exists to sharpen the boundary between /b/ which /d/, and would lead to an intensity variation as shown in Fig. 3.24 above. In Chapter 5, a model of a possible inhibitory interaction between two hypothetical neural populations is investigated, and is incorporated into the SDT model of Section 2.3. This model is shown to possess the properties necessary to account for the identification and discrimination data presented in this chapter. First, however, two more experiments are reported: selective adaptation and binaural fusion, both of which shed light on the nature of the interactions in the perception of the /b/-/d/ stimulus combinations.

CHAPTER 4

SELECTIVE ADAPTATION

From the results of Sawush and Pisoni (1976), Miller (1977) and, to a limited extent, from Ainsworth (1977), it can be inferred that the amount of boundary shift under selective adaptation is related to the position of the adaptor along the test continuum. As the adaptor takes on physical values intermediate to the endpoint stimuli on a two-stimulus continuum, the boundary shift changes from negative at one end of the continuum to positive at the other[1], and is zero when the adaptor is a boundary stimulus. In auditory adaptation studies, the change in threshold (temporary threshold shift) with adaptor intensity is well known (e.g., Ward et al., 1958), and recent evidence suggests that phonetic boundary shifts are likewise a function of the adaptor intensity (Hillenbrand, 1975; Simon, 1977; Miller et al., 1977; Sawusch, 1977). On this basis, selective adaptation to a /b/-/d/ combination ought to result in a boundary shift.

---

[1] The boundary shift is defined as the postadaptation boundary minus the preadaptation boundary. One direction can arbitrarily be called positive.

If the /b/ component of the composite stimulus selectively adapts the /b/ processor[2] and the /d/ component selectively adapts the /d/ processor, then the boundary shift should be regulated by the relative desensitization of these processors. When the /b/ and /d/ components are perceptually equal (i.e., the adapting stimulus is a boundary stimulus), no boundary shift is to be expected since the /b/ and /d/ processors will be affected equally. As α takes on values greater or less than α_{50}, the adapting stimulus rapidly acquires a unique phonetic identity. The weaker component then becomes subliminal. If the boundary shift is solely a function of the phonetic category of the adaptor, it follows that for this model, the boundary shift should be a function of α only in the boundary region (see the solid line in Fig. 4.1). On the other hand, if selective adaptation is selective at the auditory level rather than at the phonetic level (i.e., the amount of adaptation depends on the relative intensities of /b/ and /d/), a much more gradual influence of adaptor composition should be observed (see the dashed line in Fig. 4.1).

---

[2] It is not being proposed that /b/ and /d/ are processed as intact spectro-temporal patterns. But, since the total intensity of the /b/ and /d/ vary with α, each of their acoustic cues - whatever they may be - vary with α in the same fashion. Thus, within the context of the present paradigm, it is convenient to speak of a "/b/ processor" and a "/d/ processor".

4.1 EXPERIMENT 8: Selective Adaptation

The apparatus for the presentation of the stimuli was as already shown in Fig. 3.6. The presentation program was modified to play back an adaptor of pre-determined composition (α_a) for three minutes at a presentation rate of approximately 2 per second. The identification test of Experiment 1 was conducted, except that after every 11 stimulus presentations, a reinforcing adaptation period of 75 presentations was conducted. A one-second 1000 Hz tone signalled the commencement of the reinforcement.

Two pre-adaptation identification runs were conducted prior to conducting each adaptation run. The identification runs lasted approximately eight minutes each, and the adaptation run lasted approximately 35 minutes. The stimuli were presented monaurally to the right ear. Three subjects participated in this study, which involved adaptation to 11

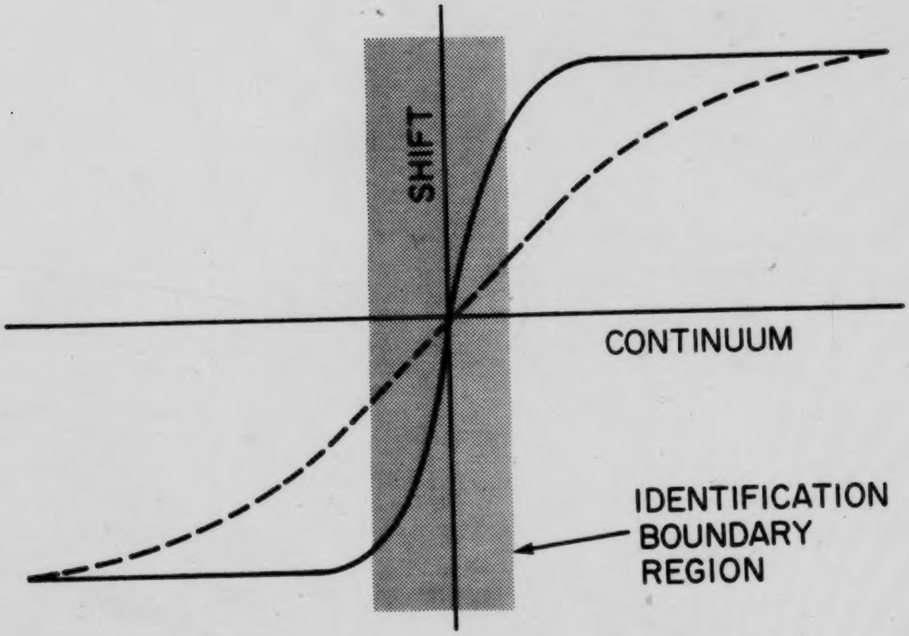


Fig. 4.1. Hypothesized boundary shifts for phonetic-level adaptation (solid line) and auditory adaptatión (dashed line). The grey region represents approximately the boundary zone

values of α_a from 0 to 1 in steps of $\Delta\alpha_a = 0.1$. The value of α_a was chosen at random for any particular run. A second run for each α_a was obtained for subject DS.

Fig. 4.2 shows a few typical pre- and post-adaptation runs. The boundary shifts were computed by fitting a normal ogive to all identification curves. The axes of Fig. 4.3 are expressed in decibels, where

$$I'_s = I'_{50} - I_{50} \tag{4-1}$$

were I'_{50} (the shifted boundary) and I_a (the adaptor) are computed according to Equation 3-8. I_{50} is the mean pre-adaptation boundary for that run, as determined from the pre-adaptation identification tests. Table 4-1 shows the pre- and post-adaptation boundaries for the three subjects.

Subjects GB and JH demonstrate very similar boundary shifts over the entire α range. Since their average pre-adapted boundaries differed by only 1.2 dB, their data were averaged.[3]

---

[3] All averages were performed with the boundary locations specified in decibels. Thus, the means are harmonic means.

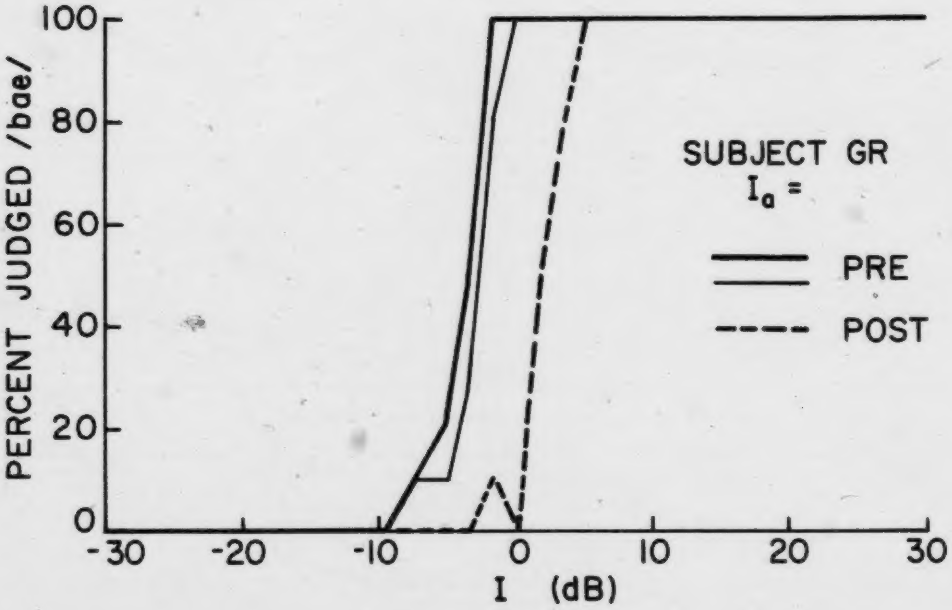


Fig. 4.2. Typical pre- and post-adaptation identification curves

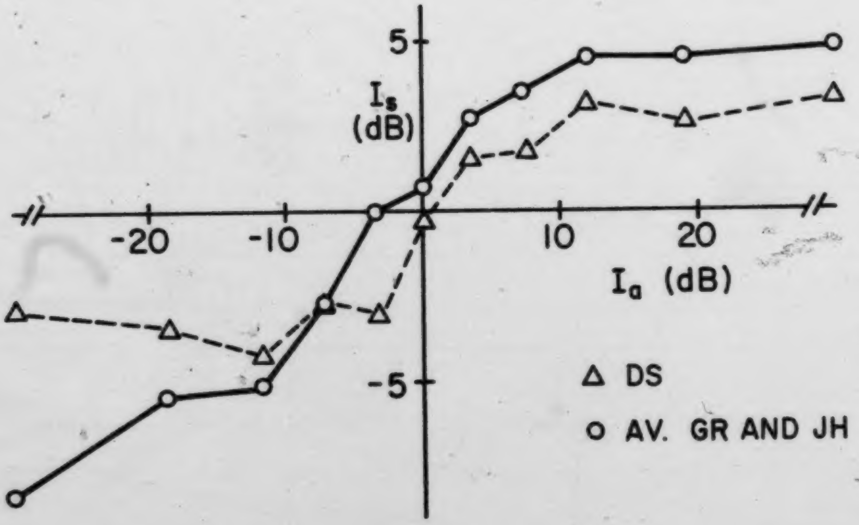


Fig. 4.3. Boundary shift, I_s, as a function of adaptor composition

TABLE 4-1

PRE- AND POST-ADAPTATION BOUNDARIES

	JH		GR		DS	
ADAPTOR	UNADAPT	ADAPT	UNADAPT	ADAPT	UNADAPT	ADAPT
I_a (dB)	I_{50} (dB)	I'_{50} (dB)	I_{50} (dB)	I'_{50} (dB)	I_{50} (dB)	I'_{50} (dB)
	-0.37	-8.36	0.17	-8.32	2.45	-0.45
-19.0	-0.34	-5.91	-0.13	-5.35	2.56	-0.86
-12.0	-1.56	-6.79	0.40	-4.60	2.72	-1.47
-7.4	-0.59	-3.53	-0.03	-2.42	2.51	-0.19
3.5	-1.40	-2.05	0.34	-0.31	3.01	-0.13
0.0	-0.87	-0.23	0.12	0.87	2.72	2.37
3.5	-0.96	2.63	-0.40	1.71	2.77	4.26
7.4	-0.60	3.39	0.88	4.72	2.83	4.46
12.0	-0.55	3.72	0.13	4.66	2.05	5.13
19.0	-1.08	3.52	0.00	4.35	3.17	5.72
	-3.35	1.98	0.15	4.38	2.66	5.82

The data for subject DS were not included in the averaging since the differences in boundary shifts between subject DS and either GR or JH appear to be of the order of a factor of two.* Fig. 4.3 shows that significant boundary shifts are produced for non-boundary adaptors. This suggests that phonetic-level adaptation is not the prime determinant of the boundary shift, since boundary shifts are a strong function of adaptor composition for adaptors up to 10 dB or so from the normal unadapted boundary. It appears that whatever physiological mechanism is reponsibile for the shift it is sensitive to the acoustic form of the signal and not just its phonetic value.

There is an observation to be made concerning the perceived quality of the stimuli during the adaptation run. Approximately one-third of the way into the test, both the adaptor and the test stimuli start to sound "fuzzy" or "rough", as if noise were being added to the signal. This is characteristic of tonal signals in under auditory fatigue (Hirsh and Ward, 1952; Davis et al., 1950) but so far has not been noted in the selective adaptation literature. It was consistent for all runs, and noticed by all three subjects. In general, test stimuli which were of the same category as the adaptor tended to sound noiser than those belonging to the opposite category. In spite of this change of stimulus quality, no discernible influence on the slope of the identification curve is observed (see Fig. 4.4). The first point on Fig. 4.4 corresponded to an identification curve with an abnormally large slope, and this point also corresponds to an abormally large boundary shift (see Fig. 4 .3). Both JH and GP exhibited large boundary shifts and large boundary regions for this particular run. Two replications (see Experiment 9) of this point by subject JH failed to reproduce either the large shift or large slope, so this point must be considered suspect).

---

* This was also observed in a pilot study in which a complete run was performed on subject DS and a few test points were obtained from subject JH. The boundary shifts for JH were a factor of two greater in this study also.

Sawusch (1977) and Miller (1975) used confidence ratings to measure within-category changes in VOT judgements due to adaptation, and their results show that the quality rating decreased for within-category stimuli. This they interpreted as a desensitization of the detector response functions over their entire domain. Although neither Miller nor Sawusch indicate the nature of this qualitative change in the percepts, it is possible that it is due in part to increased noisiness of the signal as found in the present experiment. As one possible explanation, Sawusch suggests that

> "... the within category effect seems to be characteristic of adaptation at the peripheral frequency specific auditory level and a fatigue interpretation of the adaptation at this level seems to be appropriate. " (p.749).

He also suggests as an alternative explanation a "retuning"

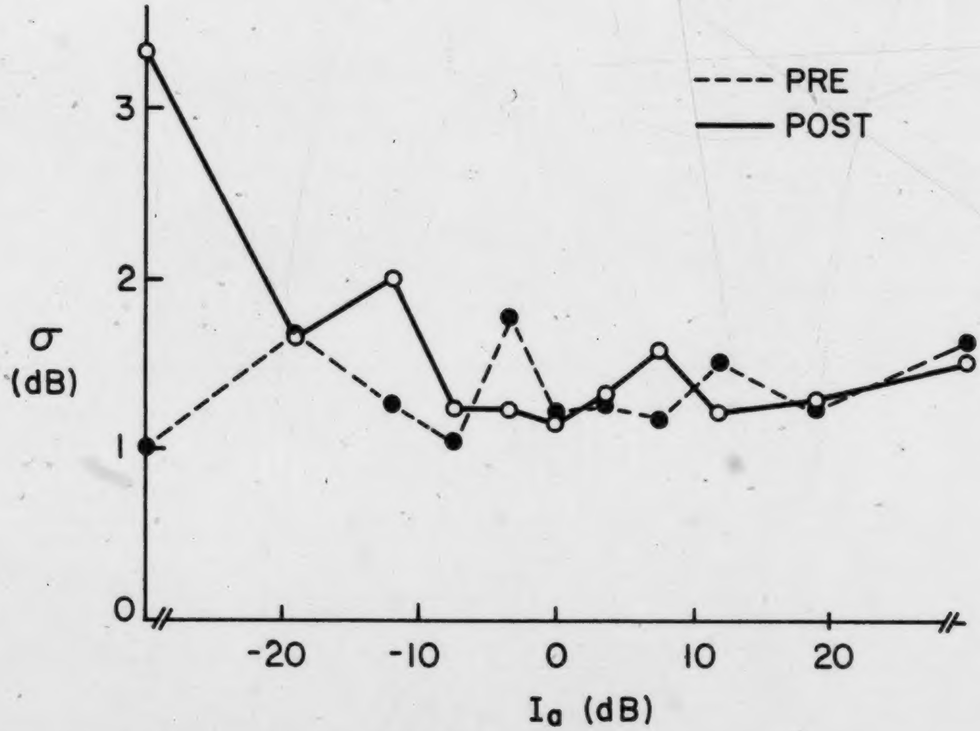


Fig. 4.4. Width of the identification boundary (σ) as a function of adaptor composition

operation at the central level, but the type of change in signal quality observed in the present experiment would argue in favour of the first interpretation.

4.2 Summary

Experiment 8 provides information on how the /b/ and /d/ processors interact when selectively adapted, and it is clear that there are within-category adaptation effects. This finding is consistent with the results of Sawusch (1977) and Miller (1975), and rules out the possibility that the adaptation is occurring at a level where only the phonetic identity of the adapting stimulus is preserved. Also, it was noted during the selective adaptation runs that repeated playback of an adaptor close to the boundary does not result in a "flip-flop" of the percept between the two phonetic categories, as may occur in reversible visual figures (Taylor and Aldridge, 1974). Rather, the adaptor is perceived as a slightly noisy exemplar of one of the categories, and this identity persists for the entire adaptation run. While it is possible to induce the opposite percept for a stimulus near the boundary, there always appears to be a preferred percept on repeated playback (see Section 3.2.6). In this light, phonetic-level adaptation is ruled out since the boundary shifts then should only take on two discrete values, one for either category of adaptor. Fig. 4.3 shows that this is not the case.

The alternate possibility (which will be explored with a mathematical model in Chapter 5) is that the degree of adaptation is strictly determined by the acoustic composition of the adaptor. If the /b/ and /d/ components are recognized by separate processors, neural adaptation at a level where the spectro-temporal information is preserved should result in a reduced input input to each processor. The outputs of these two processors ("detectors", if you will) would then be reduced by an amount which depends on the composition of the adaptor, and a boundary shift will ensue.

In order to pursue this line of analysis, more information on the response characteristics of these hypothesized processors is necessary. In particular, it is desirable to know how the boundary shift depends on the intensity of a single (i.e., pure /bae/ or /dae/) adaptor. This information is necessary to decide if the boundary shift for the composite stimulus can be predicted from the effects of each component independently.

4.3 EXPERIMENT 9: Effect of Adaptor Intensity

Experiment 8 was replicated using /dae/ adaptors of various intensities. Two of the original three subjects participated. Experimental sessions were conducted every two or three days at the subjects' convenience (it is desirable to leave at least 24 hours between adaptation sessions to avoid any cumulative effects). The adaptor intensities were chosen in the range 0 dB to -17 dB re the full /dae/ stimulus intensity. The intensity on any particular session was selected at random.

The results (Fig. 4.5) show that the boundary shift increases with adaptor intensity as expected. Significant boundary shifts commence when the /dae/ adaptor intensity is of the order of 50 dB SPL, which is consistent with threshold shifts for exposure to narrow band noise (Ward et al., 1958; Trittipoe, 1958). The boundary shifts for /dae/ have roughly the same reproducibility as those shown in Fig. 4.3, but since the shifts are everywhere smaller, the uncertainty is proportionally greater. Several points were replicated in an attempt to stabilize the boundary shift estimates. Fig. 4.6 shows the boundary shifts for the composite adaptors and the /dae/ adaptors plotted on the same graph. From this diagram, it can be seen that the rate of change of boundary with change in intensity of the /dae/ adaptor is greatest for low adaptor intensities. Assuming that the effect of a /bae/ adaptor is similar, the boundary shifts predicted on the basis of each processor being affected independently by the corresponding /b/ or /d/ component can be represented to a first approximation as the difference of the boundary shifts associated with the /bae/ and /dae/ adaptors independently. This leads to a curve of

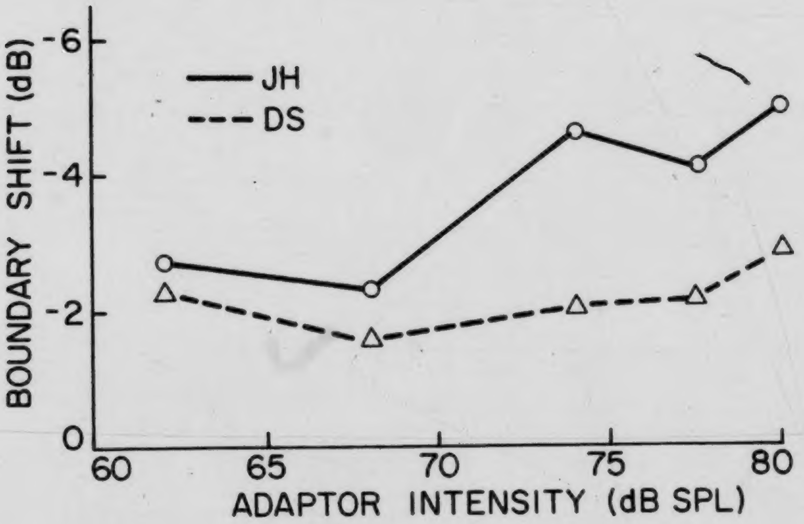


Fig. 4.5. Boundary shifts as a function of the intensity of /dae/ adaptor

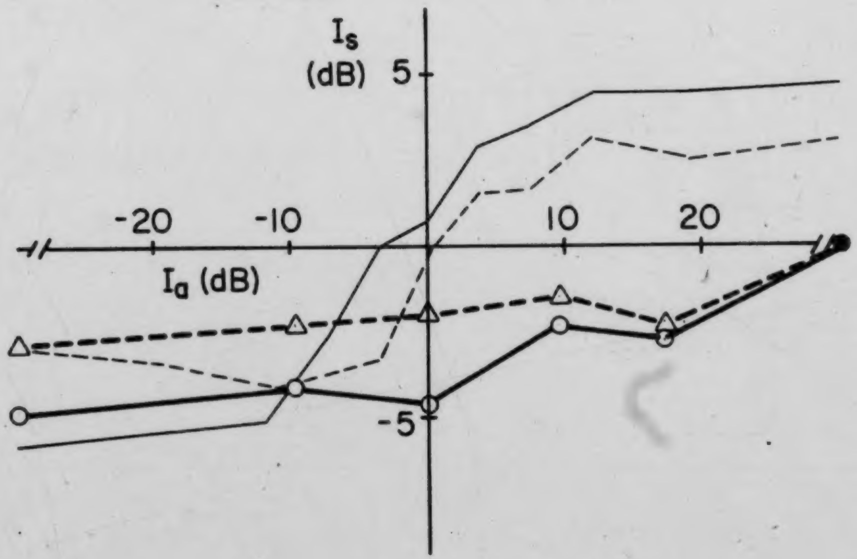


Fig. 4.6. Comparison of boundary shifts for the /dae/ adaptor (thick lines) and composite /bae/-/dae/ adaptor (thin lines)

the form shown in Fig. 4.7, and has a curvature exactly the opposite to that obtained in Fig. 4.3. Inasmuch as this is only a crude estimate of the effects of a combined /b/ and /d/ adaptor, it still suggests that the effect of the composite adaptor is not simply related to the effects of either signal component taken independently. The relationship between Experiments 8 and 9 will be considered again in Chapter 5 where a model of the adaptation process is constructed.

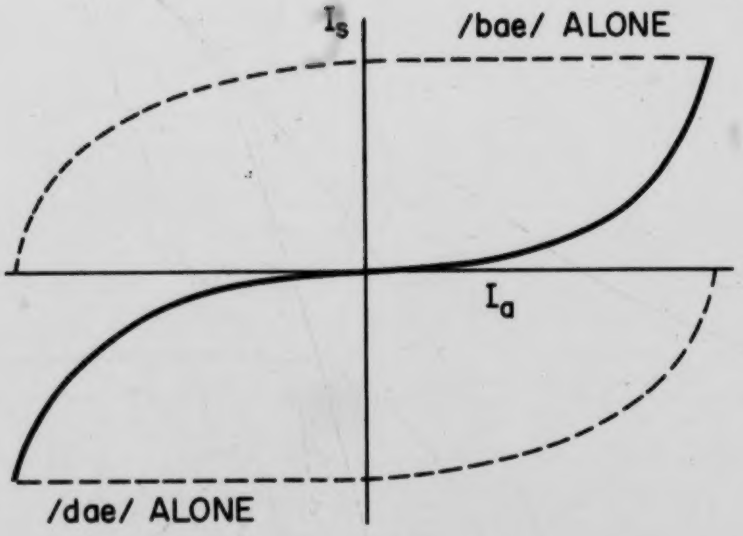


Fig. 4.7. Hypothesized boundary shifts for the case of both "pure" /bae/ and /dae/ adaptors and composite /bae/-/dae/ adaptor (assuming the shift due to the composite adaptor is basically equal to the difference between the shifts associated with the "pure" adaptors)

CHAPTER 5

MODELLING THE EXPERIMENTAL RESULTS

The model of AX discrimination developed in Chapter 2 required the specification of an arbitrary dispersion function in order to fit the experimental results. In this chapter, a model of the monaural fusion paradigm is developed which leads to the required dispersion, and is extended to include selective adaptation. The motivation for the model stems from the consideration that whatever the acoustic cues for /bae/ and /dae/ are, they must vary as the relative intensity of the respective signal components. The basic assumption is that each signal component results in a degree of excitation in a separate neural population which is monotonic with the intensity of the stimulating signal. These populations will be referred to as "detectors".

To formulate the model, a number of assumptions are necessary. The major assumption - that of separate /b/ and /d/ processors - is the most difficult to substantiate. Support for this assumption comes from the fact that a boundary stimulus can be perceived as a member of either phonetic category - thus demonstrating that both sets of acoustic cues are simultaneously available for processing. The informal test conducted during Experiment 1 to test how far the boundary could be moved by controlling overt response bias (Fig. 3.11) also supports this assumption, and shows that in the transition region the acoustic cues for both /bae/ and /dae/ are simultaneously available.

The AX discrimination results show (by virtue of the extracted dispersion function) that this relative intensity continuum is categorically perceived. The fact that the dispersion function did not come out to be a delta function (which is hardly surprising) indicates that some within-category discrimination is possible, and that the discriminability decreases to zero as either of the two signal components vanishes. The indications are (re Fig. 3.22) that the influence of the weaker signal component is detectable to about ±15 dB or so away from the phonetic boundary. The boundary shifts in the selective adaptation experiment (Experiment 8) also show that the weaker signal is an effective adaptor out to approximately ±15 dB, perhaps beyond. Taken together, these two experiments suggest that their respective interpretations do not depend on the phonetic value of the stimulus, but rather on its acoustic structure.

5.1 A PRELIMINARY MODEL OF /b/-/d/ DETECTION

The first attempt at a model will assume that each component of the composite stimulus is detected by a separate neural population, and that the degree of excitation in these populations scales with stimulus intensity as a power law function, as for the growth of loudness. In the discussion which ensues, it is important to note that this modelling is intended only as functional, and does not purport to have any anatomical or physiological correspondences. To commence the model, let u_1 and u_2 be variables representing the degree of neural excitation corresponding to the /d/ and /b/ components of the stimulus respectively. The excitation of the two neural ensembles can then be expressed as

$$u_1 = kI_2 \qquad (5\text{-}1a)$$

$$u_1 = kI_2 \qquad (5\text{-}1b)$$

where I and I are the intensities of the /b/ and /d/ components respectively. "k" is a normalizing constant whose value is of no interest here. Using Equations 3.4 to define the signal energies[1] of the /b/ and /d/ components, the excitation can be expressed as a function of α :

$$u_1 = k'(1-\alpha)^{2\theta} \qquad (5\text{-}2a)$$

$$u_2 = k'\,\alpha^{2\theta} \qquad (5\text{-}2b)$$

where k' has absorbed both k and the absolute conversion factor between I and α^2. Since k' merely determines the scale for u_1 and u_2, for computational purposes it can be taken as unity.

To complete this model, it is only necessary to specify the exponent θ . The commonly accepted exponent for the growth of loudness is $\theta = 0.27$ (Stevens, 1971; Zwislocki, 1969; Luce, 1977). Using $\theta = 0.27$, and Equations 5-2, u≠ and uʟ can be calculated for values of θ between 0 and 1. The result is a "stimulus trajectory" in the u_1-u_2 plane (see Fig. 5.1). Assuming, as for the detector model in Section 2.3, that u_1 and u_2 are the means of two

---

[1] Since it is not clear which energies are being monitored, the total energy (Equations 3-4) is used as an estimate of the stimulus intensity. Again, since the intensity of any cue varies as the total energy, this should not be a significant source of error.

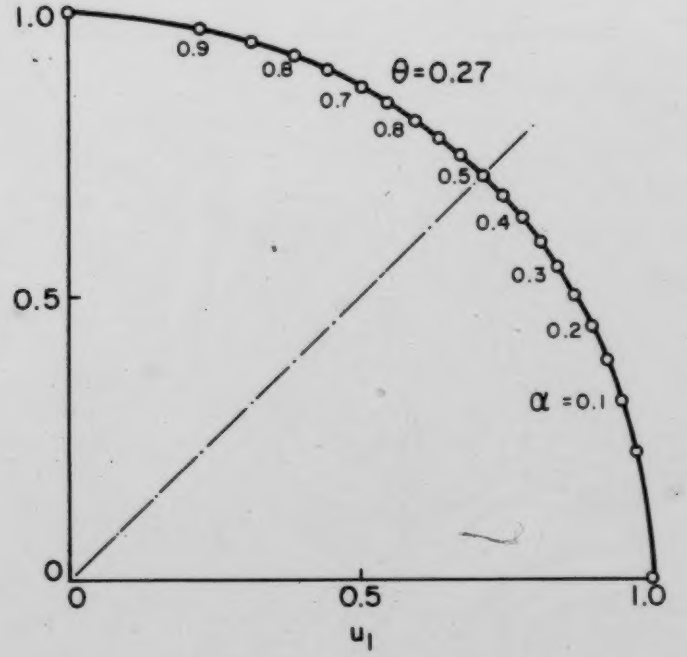


Fig. 5.1. Stimulus trajectory for θ=0.27

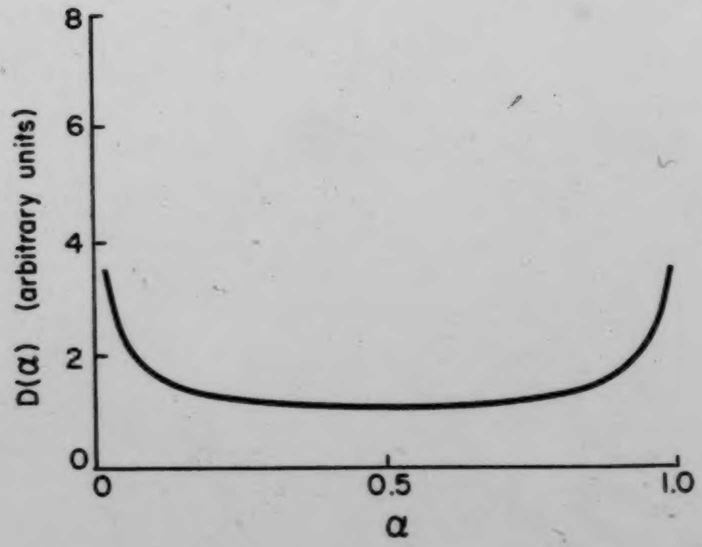


Fig. 5.2. Dispersion function for θ=0.27

(uncorrelated) random variables U_1 and U_2 with variance σ , the point (u_1, u_2) represents the centroid of a bivariate circular normal probability distribution in the u_1-u_2 plane. Using the decision variable $(U_1 - U_2)/\sqrt{2}$, the probability of identifying a given stimulus characterized by α as, say, /b/ is

$$p(\alpha) = \Phi\left[\frac{u_2(\alpha) - u_1(\alpha)}{\sqrt{2}\ \sigma}\right] \tag{5-3}$$

This is a sigmoidal curve, and thus for a suitable choice of σ will resemble the identification functions of Experiment 1. The identification data alone, however, are insufficient to test the model, since any stimulus trajectory which crosses the decision line $(u_2 = u_1)$ only once will lead to a sigmoid-shaped identification function. A more stringent test of the model is the AX discrimination data of Section 3.4. From Section 3.3., a suitable dispersion function was determined to be

$$D(\alpha) = \frac{u_1(\alpha)\ u_2'(\alpha) - u_1'(\alpha)\ u_2(\alpha)}{u_1^2(\alpha) + u_2^2(\alpha)} \tag{5-4}$$

Using u_1 and u_2 defined by Equations 5-2 above, this becomes

$$D(\alpha) = \frac{2\theta[(1-\alpha)^{2\theta}\alpha^{2\theta-1} + (1-\alpha)^{2\theta-1}\alpha^{2\theta}}{(1-\alpha)^{4\theta} + \alpha^{4\theta}} \qquad (5\text{-}5)$$

For $\theta = 0.27$, the dispersion function appears as shown in Fig. 5.2. This dispersion function shows no enhanced dispersion, and an attempt to fit this model to the discrimination results would be fruitless.

Equations 5-2 and 5-3 constitute a model with two parameters: θ and σ. Now, it should be clear from Section 3.2 that if the dispersion function is approximately constant, changing σ will not generate enhanced dispersion. The only possiblity that remains is to allow to take on different values. Fig. 5.3 shows various stimulus trajectories and corresponding dispersion curves for $\theta = 0.27$ to $\theta = 2.0$. It can be seen that as becomes large, the dispersion function does indeed become unimodal, and by making θ arbitrarily large, the peak of the dispersion function can be made arbitrarily narrow. It would appear, then, that in order to make this model fit the discrimination data, it is necessary to let $\theta = 2.0$ or greater. However, this seems unacceptable since such large exponents are not noted for the growth of loudness with stimulus intensity. As will be shown below, inclusion of some form of inhibition between these hypothetical neural populations will produce the required dispersion. In order to do this, however, it is first necessary to formulate a more detailed model of the excitation process.

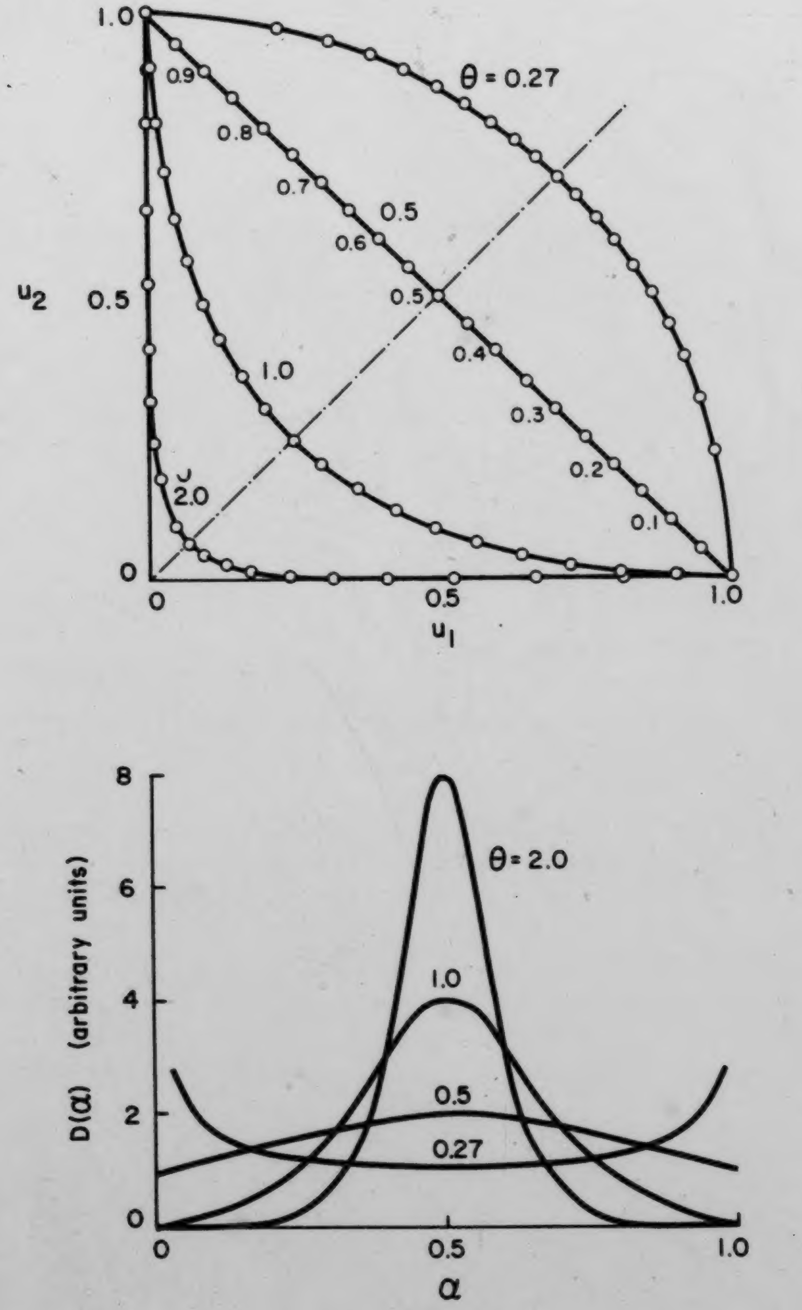


Fig. 5.3. (a) strimulus trajectories and (b) dispersion functions for various values of θ

5.2 INTENSITY CODING BY NEURAL POPULATIONS

Consider a population consisting of N_0 hypothetical neurons. For the sake of the present discussion, these cells are assumed to be capable of becoming excited (i.e., firing) in response to external stimulation. N_0 is assumed sufficiently large that the mean level of excitation is insensitive to minor perturbations by individual neurons in the population. The fundamental assumptions are:

(a) the probability of firing is the same for all cells in the population which are in the "ground state"

(b) the probability per unit time of a cell firing in response to a stimulus of intensity I is given by

$$\gamma' = \gamma\, f(I) \qquad (5\text{-}6)$$

where f(I) initially will be assumed to be a power law, i.e., $f(I) = I^{\theta}$.

Now, a cell, once excited, is no longer capable of firing for a certain time (i.e., it is in a refractory state). After a time typically of the order of 1 or 2 milliseconds, the cell recovers, and the probability of recovery per unit time will be denoted by γ_r . Zwislocki (1969), for a neurally-based model of temporal summation, suggests a value of 300 $\sec^{-1}$ for $\gamma_r = 1/\tau_r$, and this value will be assumed here. The process of recovery will be referred to as "de-excitation". The excitation/de-excitation process is illustrated schematically in Fig. 5.4.

The rate processes shown in Fig. 5.4 lead to the differential equation

$$\frac{dN}{dt} = \gamma f(I)N_g - \gamma_r N \qquad (5\text{-}7)$$

where N is the number of cells in the ground state at time t, and N is the number of excited cells. Requiring that the total number of cells in the population remain constant, i.e.,

$$N_0 = N + N_g \qquad (5\text{-}8)$$

Equation 5-7 becomes

$$\frac{dN}{dt} = \gamma f(I)N_0 - (\gamma f(I) + \gamma_r)N \qquad (5\text{-}9)$$

As long as I is invariant with time, this differential equation has constant coefficients, and, assuming the

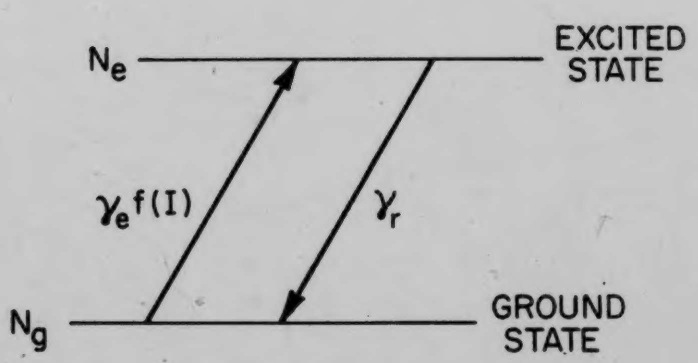


Fig. 5.4. Schematic model of excitation/de-excitation processes in a hypothetical neural population

initial condition $N(0) = 0$, has the solution

$$N(t) = \frac{\gamma f(I)N_0}{\gamma f(I)+\gamma_r}\left[1-e^{-(\gamma f(I)+\gamma_r)t}\right] \qquad (5\text{-}10)$$

N_0 can be arbitrarily taken as unity, in which case $N(t)$ then represents the proportion of excited cells in the ensemble.

This result shows that a population of cells can behave as an energy integrator with a time constant given by

$$\tau = \frac{1}{\gamma f(I)+\gamma_r} \qquad (5\text{-}11)$$

From Equation 5-11, it follows that if the probability of a cell firing increases with stimulus intensity, the time constant of the integration decreases with intensity. Such a reduction in time constant with stimulus intensity has been noted many times in auditory and visual temporal summation studies (cf Roufs, 1975; Marks, 1972; Stevens and Hall, 1966; Small, Brandt and Cox, 1962; Miller, 1948). For times $t \ll \tau$, Equation 5-10 can be approximated by

$$N(t) \sim N_0 \gamma f(I)t \qquad (5\text{-}12)$$

which, assuming $f(I) = I^{\theta}$, is the familiar "law" of temporal summation (i.e., for constant N , $I^{\theta}t$ = constant). For longer times, $t \gg \tau$, the saturation level of excitation is just

$$N\infty = \frac{\gamma f(I) N_0}{\gamma f(I) + \gamma_r} = N_0 \frac{I^\theta}{I^\theta + I_0^\theta} \qquad (5\text{-}13)$$

where $I_0^\theta = \frac{\gamma_r}{\gamma}$. This particular response function has enjoyed a great deal of popularity in the vision literature (e.g., Mansfield, 1976; Marks, 1972, 1974; Alpern, 1971) and has also been investigated as a model of the transfer function of sensory transducers (Lipetz, 1971).

For intensities I such that $I^\theta << I_0^\theta$, Equation 5-13 reduces to the familiar psychophysical power law:

$$N\infty \sim \frac{I^\theta}{I_0^\theta} \qquad (5\text{-}14)$$

Thus, for intensities low enough that saturation does not occur in the neural mass, this model is linear with respect to the driving function f(I) (I^θ in this case). Any nonlinearity with respect to I comes from the driving function. Therefore, in order to obtain power law behaviour with this model, it is necessary to supply it as f(I).

Equation 5-11 , since it yields a power law behaviour for non-saturating intensities, is not an improvement over Equations 5-1. Admittedly, this solution provides the time dependence of the excitation process, but only the steady-state solution (5-13) is of interest here. Although it is hardly realistic to assume that I = constant, since it is not known exactly which energies in the /b/-/d/ stimuli are being monitored, there is little point in dwelling on the temporal behaviour of N . Also, from a practical standpoint, if stimulus intensity were not considered independent of time, Equation 5-9 may or may not have an analytic solution. In any event, this formulation of the model is only an intermediate step to enable the inclusion of mutual inhibition between the two neural populations.

5.3 INHIBITION BETWEEN NEURAL POPULATIONS

Neural models which incorporate inhibitory processes demonstrate behaviour quite different from models which incorporate only excitation processes (Wilson and Cowan, 1972). The Wilson and Cowan model differs somewhat in its development from the one given here, but is based on similar reasoning and premises, and demonstrates similar behaviour.

Consider the two neural populations shown schematically in Fig. 5.5. Suppose that the degree of activation of the inhibitory connections is some function of the excitation (N_1) in the parent population (i.e., the population doing the inhibiting). The probability per unit time that a cell in population 2 will be inhibited can then then be expressed as $\gamma_{i2}\, h(N_1)$, where N_1 is the level of excitation in the first population, and h is some as yet unspecified function. The excitation in the two masses is then governed by the following four differential equations

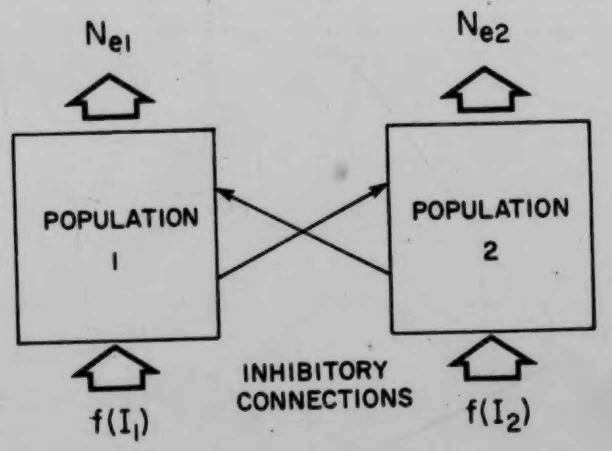


Fig. 5.5. Schematic configuration of two mutually inhibiting neural populations

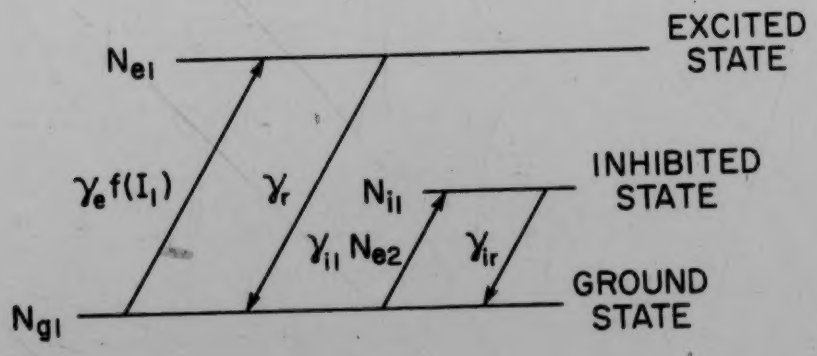


Fig. 5.6. State diagram of two mutually inhibiting neural populations

$$\frac{dN_1}{dt} = \gamma f(I_1)N_{g1}-\gamma_r N_1 \tag{5-15a}$$

$$\frac{dN_{i1}}{dt} = \gamma_{i1}h(N_2)N_{g1}-\gamma_{ir}N_{i1} \tag{5-15b}$$

$$\frac{dN_2}{dt} = \gamma f(I_2)N_{g2}-\gamma_r N_2 \tag{5-15c}$$

$$\frac{dN_{i2}}{dt} = \gamma_{i2}h(N_1)N_{g2}-\gamma_{ir}N_{i2} \tag{5-15d}$$

These equations can be represented by the state-level diagram shown in Fig. 5.6. In this model, a neuron can either be excited, inhibited, or in the ground state, and furthermore, can only be in one of these states at one time. Equations 5-15 are typical of Volterra-style population models, and techniques for investigating their possible solutions and stability points have long been a topic of interest to mathematicians and biologists. In the present case, no direct significance can be attached to the time dependent solutions, so only the steady-state solution will be considered. This is obtained by setting the above derivatives to zero, i.e.,

$$\gamma f(I_1)N_{g1}-\gamma_r Ne_1 = 0 \tag{5-16a}$$

$$\gamma_{i1}h(N_2)N_{g1}-\gamma_{ir}N_{i1} = 0 \tag{5-16b}$$

(It is necessary only to show two of the equations; the two equations for the other population can be obtained by interchanging subscripts 1 and 2). Substituting $N_{g1} = N_{01} - N_1 - N_{i1}$ in Equations 5-16a and 5-16b, these become

$$\gamma f(I_1)N_{01} - (\gamma f(I_1)+\gamma_r)N_1 - \gamma f(I_1)N_{i1} = 0 \qquad (5\text{-}17a)$$

$$\gamma_{i1}h(N_2)N_{01} - \gamma_{i1}h(N_2)N_1 - (\gamma_{i1}h(N_2)+\gamma_{ir})N_{i1} = 0 \qquad (5\text{-}17b)$$

Letting $r = \gamma/\gamma_r$ and $r_{i1} = \gamma_{i1}/\gamma_{ir}$, Equations 5-17 can be rewritten to yield

$$rf(I_1)N_{01} - (rf(I_1)+1)N_1 - rf(I_1)N_{i1} = 0 \qquad (5\text{-}18a)$$

$$r_{i1}h(N_2)N_{01} - r_{i1}h(N_2)N_1 - (r_{i1}h(N_2)+1)N_{i1} = 0 \qquad (5\text{-}18b)$$

These two equations (and the two obtained by reversing the subscripts) cannot be solved explicitly for N_1 and N_2, but can be condensed to the following implicit relations:

$$N_1 = \frac{rf(I_1)N_{01}}{rf(I_1)+1+r_{i1}h(N_2)} \qquad (5\text{-}19a)$$

$$N_2 = \frac{rf(I_2)N_{02}}{rf(I_2)+1+r_{i1}h(N_1)} \qquad (5\text{-}19b)$$

Note that when $r_{i1} = r_{i2} = 0$ (i.e., no inhibition), these reduce to the steady state solution found previously for the non-inhibitory case (Equation 5-13).

N_1 and N_2 can only be calculated from Equations 5-19 if the function $h(N)$ is specified. There is no obvious choice for h, so the simplest solution is to choose a class of functions which includes the trivial function $h(N) = N$ as one of its members. A suitable general function (which has no particular significance except that it is monotonic) is

$$h(N) = N^n \tag{5-20}$$

Using $f(I)=I^\theta$ together with $I= \alpha^2$, N_1 and N_2 are then defined by

$$N_1 = \frac{r(1-\alpha)^{2\theta}}{r(1-\alpha)^{2\theta}+1+r_{i1}N_2^n} \tag{5-21a}$$

$$N_2 = \frac{r\alpha^{2\theta}}{r\alpha^{2\theta}+1+r_{i2}N_1^n} \tag{5-21b}$$

For n=1, these equations can be solved explicitly for N_1 and N_2, but not for powers greater than 1. However, the solution for N_1 and N_2 can be iteratively computed using Newton's method (e.g., Conte, 1964, p.43).

Fig. 5.7a shows typical isoclines for n=1 and n=2 for $\alpha = 0.4$, and Fig. 5.7b shows the corresponding stimulus trajectories. The n=1 model shows a stimulus trajectory reminiscent of that produced for large values of θ in the

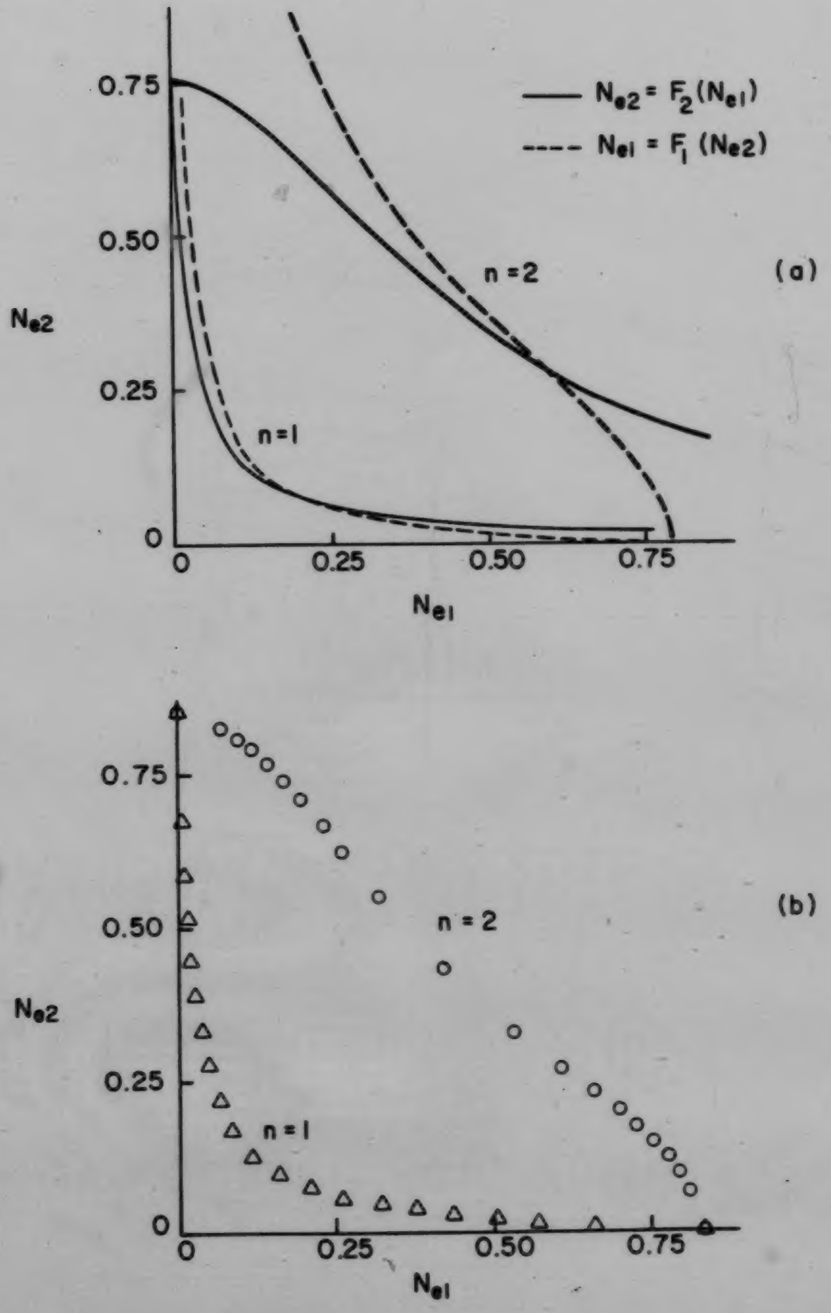


Fig. 5.7. (a) isoclines for α=0.4 for n=1 (thin lines) and n=2 (thick lines). (b) stimulus trajectories for the n=1 (triangles) and n=2 (circles) models

non-inhibitory model (Fig. 5.3b), and is undesirable for the same reasons. The n=2 model, on the other hand, shows genuine dispersion in the N_1 -N_2 plane. The reason for the dispersive power of the n=2 model can be seen by considering how N(α)grows with α for the two populations. These curves are shown in Fig. 5.8a for n=1 and in Fig. 5.9a for n=2. and the corresponding dispersion curves are shown in Fig. 5 .8b. The n=2 model achieves its dispersion by sharpening the difference between the excitation levels N_1 and N_2 (evidenced by the inflection in the N vs. α curves).

The dispersion curves shown in Fig. 5.8b show an undesirable property: the dispersion becomes infinite for α =0 and α =1. This occurs because $f(I) = I^{\theta}$ has an infinite slope at I = 0 when θ is less than unity. This will create problems for the model in trying to accommodate the data for small or large values of α , and a correction factor may be required to correct this defect. A possible correction is presented later.

This is the basic model for identification and discrimination. A given stimulus of composition specified by α results in two levels of excitation, N_1 and N_2 . All judgements concerning the categorization and discrimination presumably are transformations of these two variables. Using the mechanics already established in Section 3.3 for dispersion of a two-detector configuration, the application

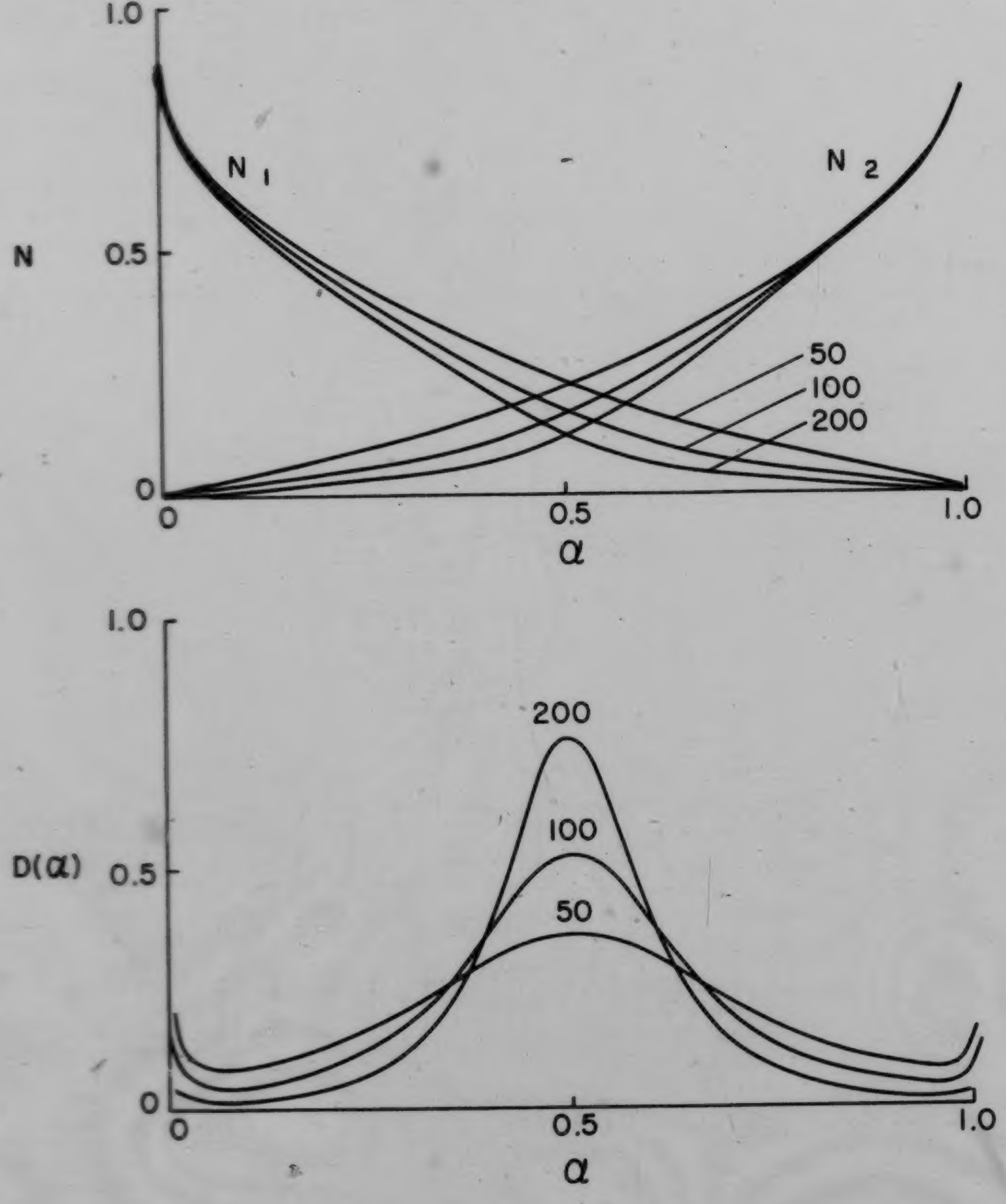


Fig. 5.8. (a) excitation curves $N_1(\alpha)$ and $N_2(\alpha)$ for the n=1 model. The curves differ by values of r_1, the degree of mutual inhibition. (b) dispersion curves corresponding to the excitation curves shown in (a)

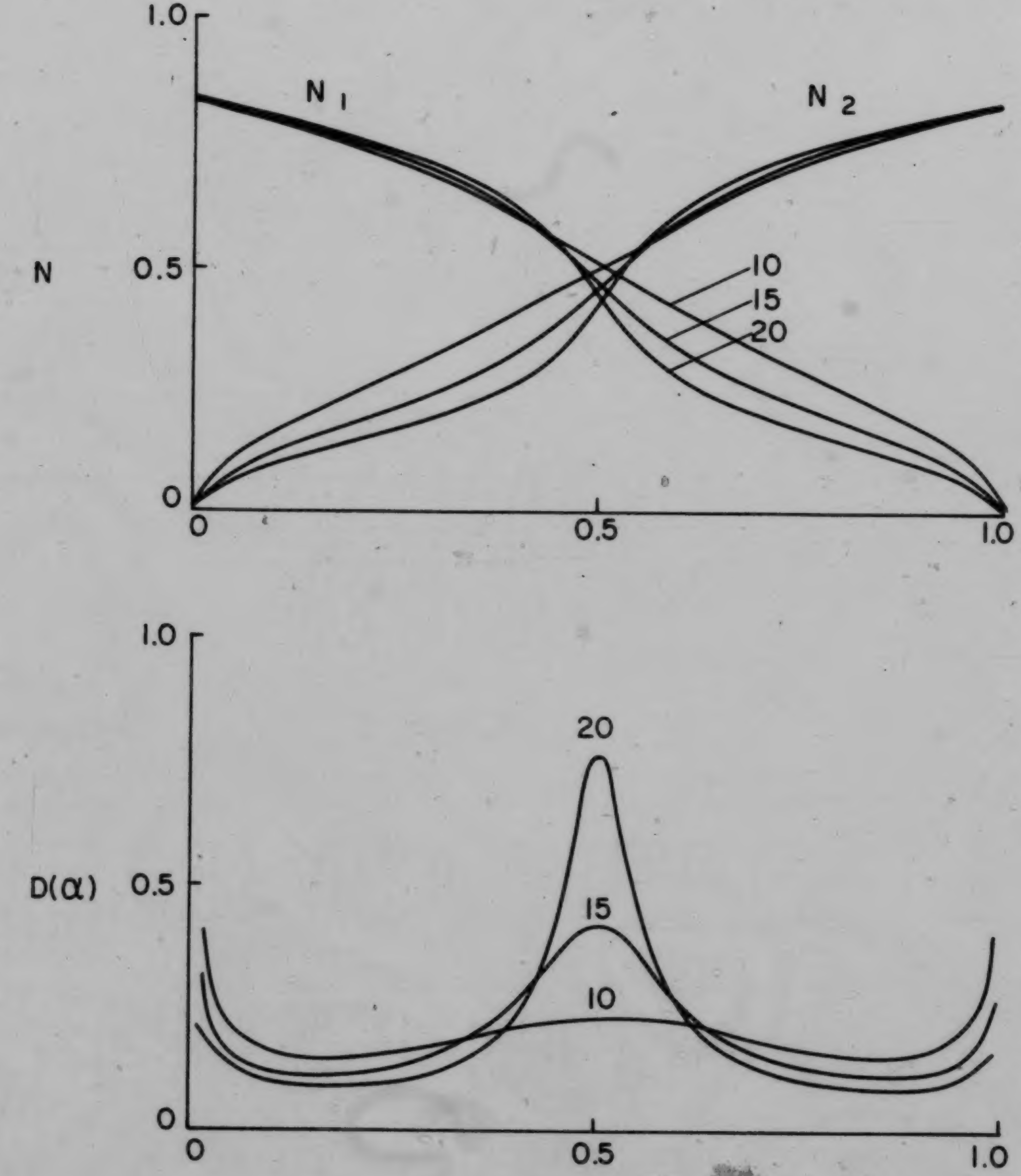


Fig. 5.9. (a) excitation curves $N_1(\alpha)$ and $N_2(\alpha)$ for the n=2 model. The curves differ by values of r_i, the degree of mutual inhibition. (b) dispersion curves corresponding to the excitation curves in (a)

of the model to the discrimination data can now proceed.

5.4 FITTING THE AX DISCRIMINTION DATA

To complete the AX discrimination model, the excitation levels N_1 and N_2 are identified with the former variables u_1 and u_2 respectively (i.e., Equations 5-2). Since N_1 and N_2 are calculated from deterministic differential equations, the randomness must be superimposed afterwards. That is, N_1 and N_2 are assumed to be the means of two random variables with equal variance σ . This assumption of constant variance may be unjustified (Luce, 1977), but a specific functional relationship for σ with intensity is lacking.[2] Any simple monotonic dependence of σ on α will probably be obscured by the flexibility of the model, so as a first approximation, σ will be assumed constant. In any event, due to the complexity of the model, simplifying assumptions are desirable until the model has been adequately investigated.

It was pointed out earlier that a plausible value for is 300 sec^{-1}. Now, the quantity $\gamma f(I)$ is also a probability per unit time, and ought to have a similar value. A nominal value of $r = \gamma/\gamma_r = 5$ is chosen, although it turns out that the the behaviour of the model is not particularly sensitive to the specific value of r. To reduce the number of free parameters in the model, r is left set at this nominal value.

---

[2] Durlach and Braida (1969), in a model for intensity detection and discrimination, make the assumption that the variance is independent of intensity. For simplicity, essentially the same assumption is made here, except that the variance is assumed to be independent of the level of excitation (which in this model is the neural counterpart of intensity).

σ is a parameter whose value is equally hard to establish since it depends on the experimental circumstances as well as the particular stimuli used. Because of this, it will be allowed to float as a free parameter in the fitting of the model, at least to obtain a representative value. It ought to be reminiscent of the difference limen for intensity, i.e., perhaps 1 dB or so (which translates on the scale to approximately 0.06). Given quantization noise, amplifier noise and recording noise, it is conceivable that it could be even larger.

Equations 5-22 now contain only two free parameters, σ and r_i. (For the present it will be assumed that $r_{i1} = r_{i2} = r_i$). Following the procedure outlined previously in Section 3.3, the AX discrimination scores were fitted by calculating the decision variable

$$y = \tan^{-1}\left(\frac{N_2-N_1}{N_1+N_2}\right) \qquad (5\text{-}22)$$

(y is the mapping corresponding to the dispersion function given by Equation 5-4, and represents the angular distance of the point (N_1 ,N_2) from the decision line shown in Fig. 2.13). An adaptive least squares fit was performed to the data shown in Fig. 3.20, and representative results are shown in Fig. 5.10. The fit is observed to be similar to that of the Gaussian dispersion function fitted in Section 3.4.1, which is to be expected since both dispersion functions have similar shapes. The influence of the enhanced dispersion for small and large values of   is evident in Figs. 5.10b and 5.10c. The failure of the dispersion curve to asymptote to zero at the ends of the scale causes the discrimination curve in Fig. 5.10c to increase rapidly away from $\alpha = 1$. This is more pronounced for the n=2 model than for the n=1 model.

The extracted values of r_i for the n=1 and n=2 models are 186 and 25 respectively. The standard deviation of the noise distribution in both cases stabilizes around 0.1 ±0.05. This value of σ is equivalent to a noise width of approximately 3 dB, which is surprisingly large since it is of the order of the width of the identification function itself. Part of the reason for this large value is that the model does not follow the trend of the data perfectly. Consequently, in the least squares fitting process, the parameters adopt whichever values are necessary to minimize the overall sum of squares difference, and this can be

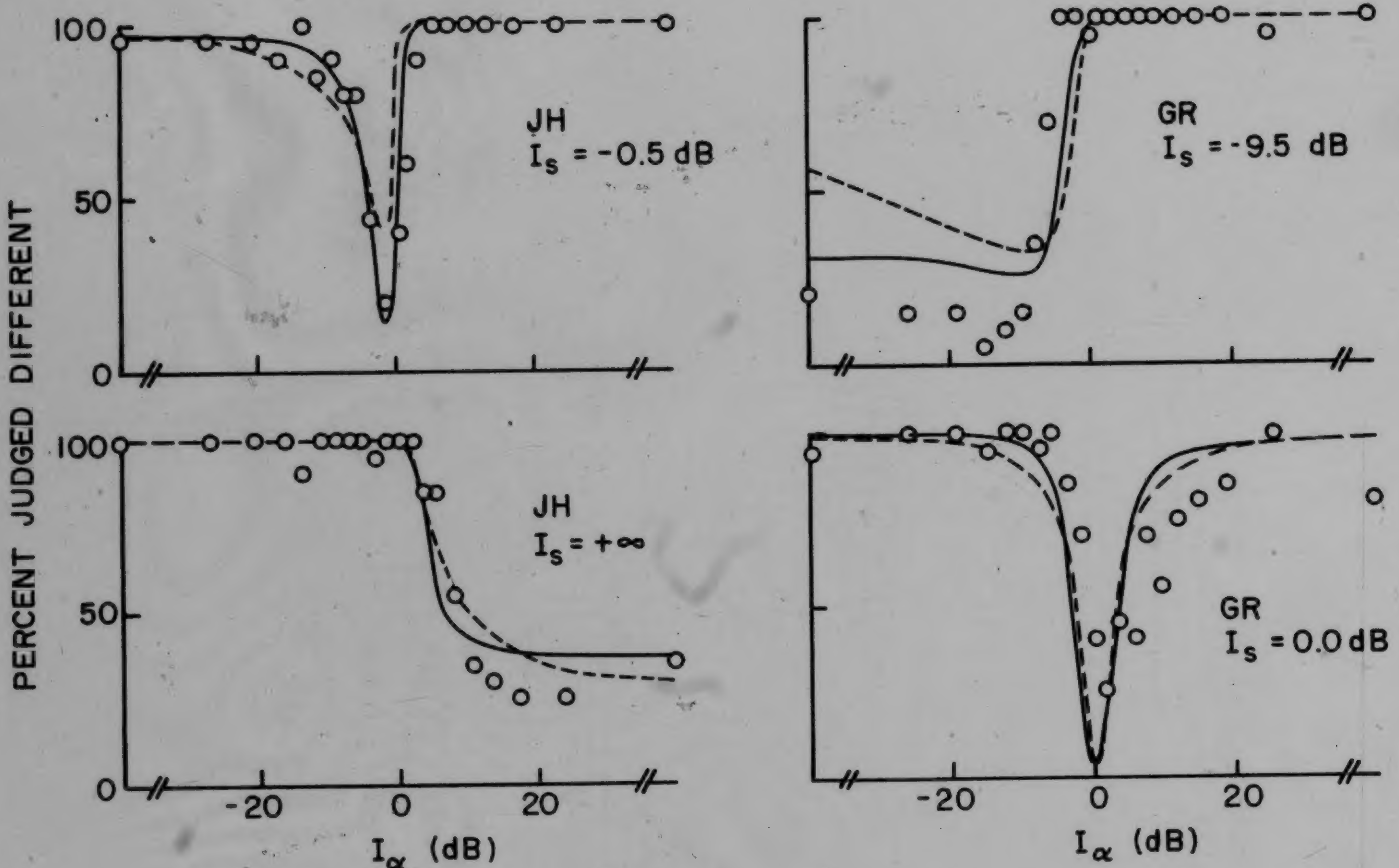


Fig. 5.10. Comparisons of n=1 (solid lines) and n=2 (dashed lines) models for AX discrimination data

accomplished in many ways. Also, since each AX discrimination profile is fitted with its own bias factor (see Section 3.4), there is not necessarily a unique minimum sum of squares. More importantly, however, there is a tradeoff between σ and r_i. The effective dispersion function can be roughly approximated by a convolution of the noise distribution with the actual dispersion function, and therefore an increase in σ can be approximately accounted for by an increase in r_i. This tradeoff prevents accurate determination of these two parameters. Nonetheless, the values of σ and r_i quoted above will be used in the selective adaptation model derived below, since their combined effect produces the required dispersion.

The n=1 model provides a superior fit for the AX discrimination data, mostly due to the more acceptable low intensity behaviour of the dispersion function (compare Fig. 5.8b with Fig. 5.9b). In the following section, the above model of the /b/-/d/ recognition paradigm is extended to include the effects of the selective adaptation experiments (Experiments 8 and 9), and it will be seen at that time that the n=1 model cannot account for these data. But, it must be remembered that the choice of inhibition function (Equation 5-21) is arbitrary, so selecting between these two models on the basis of superiority of fit does not verify either choice of model. The present interest is only in demonstrating the basic functional form of the model, since this model is only one member of a class of models.

5.5 MODELLING SELECTIVE ADAPATION

Auditory threshold shifts due to continued exposure to tones or noises have been investigated since Hood (1950). Although it has never been decisively established that adaptation/fatigue ~~results~~ in a loudness decrement per se (cf Petty, Fraser and Elliott, 1970), this is one of the common explanations (Small, 1963; Hood, 1950). Whatever the origin of adaptation/fatigue, certain regularities are observed. First, the amount of threshold shift increases with the duration of the adaptor (Ward, Glorig and Sklar, 1958, 1959). Second, it increases with the intensity of the adaptor, at least for adapting intensities up to 110 dB SPL or so (Trittipoe, 1958; Ward et al., 1958; Selters, 1964). Third, recovery is more or less exponential, with several components with different time constants being identifiable (Botsford, 1968; Hirsh and Ward, 1952). Several of these components have short time constants, possibly representing some form of neural adaptation or renewable metabolic processes. One of the components has a time constant of the order of hours, perhaps representing auditory fatigue (Botsford, 1968).

Selective adaptation studies use a similar experimental paradigm but use "phonetic boundary shift" as the measured variable. The phonetic boundary shifts in selective adaptation studies are observed to (a) increase with adaptor intensity (Sawusch, 1977; Miller, Eimas and Root, 1977; Experiment 9, this thesis), and (b) increase with number of adaptor presentations (Simon and Studdert-Kennedy, 1978). Furthermore, the phonetic boundary evidently returns to normal or near-normal within a few hours or so, which indicates that the adaptation effects are due to renewable physiological processes. While there is little agreement in either experimental domain concerning the locus or origin of adaptation/fatigue, the effects themselves are nonetheless real and also fairly reproducible.

While it is not likely to be the case that all of the TTS or phonetic boundary shift is due to neural adaptation, it seems reasonable to suppose that some of it is, especially for low intensity adaptors. In any event, this assumption will be made for the present modelling purposes since it allows a straightforward inclusion of the effects of neural adaptation into the model derived in the previous section. To incoporate adaptation as an aspect of the excitation of a neural population, an "adaptation level" is added to the state diagram of Fig. 5.6. The state-level diagram then appears as shown in Fig. 5.11. This new level represents a level at which cells may "collect", thus removing them from possible further excitation. The return to the ground state provides for the observation that

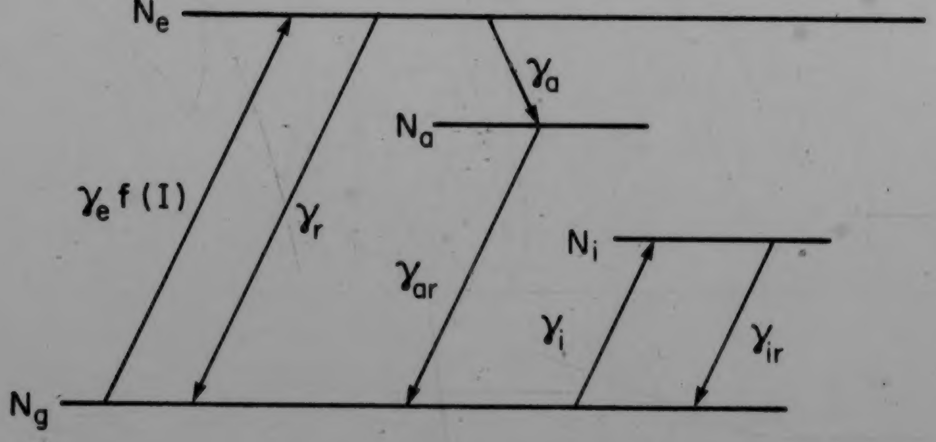


Fig. 5.11. State diagram including both inhibition and adaptation (symbolized by level N_a)

adaptation effects are temporary and wear off with the passage of time. Within the context of this model, when stimulation ceases, all adapted states will eventually return to the ground state at a rate determined by γ_{ar}. Since the data to be analyzed do not involve the temporal course of entry into or recovery from adaptation/fatigue, only one component (represented by the single level in Fig. 5.11) will be considered. A more general model would include perhaps four levels, each representing a different time constant, but for the present purposes, only one will be considered.

This appears to be a reasonable first approximation to the adaptation process. When stimulation begins, the number of cells in the adapted state will be nil, but as stimulation continues, the number of "adapted cells" will increase. For prolonged or intense stimulation, the number of fatigued cells may become a significant fraction of the total number of cells in the population. Beause of the return path to the ground state, the number of adapted cells will stabilize for a given intensity of adapting stimulation. This is consistent with experimental data since changes of threshold with adaptation and recovery are approximately exponential (e.g., Wright, 1959; Keeler, 1968). Evidence for the existence of simultaneous but independent fatigue and recovery processes is provided by Ward, Glorig and Selters (1960). In this experiment, subjects were first subjected to 105 dB SPL octave-band noise (1200 - 2400 Hz) for thirty minutes, followed by exposure to the same noise at 95 dB SPL. The TTS curves show that following the reduction in the intensity of the fatiguing signal, a decrease in TTS occurs, followed by a gradual increase (see Ward et al., 1960, Fig. 2). They conclude that this cannot be explained in terms of rate processes (charging and discharging of a capacitor in their analogy), but the Keeler (1968) analysis based on two time constants rather than one shows that this behaviour is indeed possible with such a model. In all liklihood, if the phenomenon of TTS is characterized by rate processes, the rates must be intensity-dependent.

The adaptation model is derived by generalizing rate equations 5-15:

$$\frac{dN_1}{dt} = \gamma f(I_1)N_{g1} - \gamma_r N_1 - \gamma_a N_1 \qquad (5\text{-}23a)$$

$$\frac{dN_{a1}}{dt} = \gamma_a N_1 - \gamma_{ar} N_{a1} \qquad (5\text{-}23b)$$

$$\frac{dN_{i1}}{dt} = \gamma_{i1} h(N_2) N_{g1} - \gamma_{ir} N_{i1} \qquad (5\text{-}23c)$$

Using the constraint that $N_0 = N_{g1} + N \ + N_{a1} + N_{i1}$, Equations 5-24 become (on setting the derivatives to zero for the steady state):

$$\gamma f(I_1)N_{01}-(\gamma f(I_1)+\gamma_r+\gamma_a)N_1-\gamma f(I_1)N_{a1} = 0 \qquad (5\text{-}24a)$$

$$\gamma_a N_1-\gamma_{ar}N_{a1} = 0 \qquad (5\text{-}24b)$$

$$\gamma_{i1}h(N_2)N_{01}-\gamma_{i1}h(N_2)N_1-(\gamma_{i1}h(N_2)+\gamma_{ir})N_{i1} = 0 \qquad (5\text{-}24c)$$

Dividing Equation 5-25a by γ_r, 5-25b by γ_{ar} and 5-25c by γ_{ir}, these equations become

$$rf(I_1)N_{01}-(rf(I_1)+1+r_a)N_1-rf(I_1)N_{a1} = 0 \qquad (5\text{-}25a)$$

$$r_{ar}N_1-N_{a1} = 0 \qquad (5\text{-}25b)$$

$$r_{i1}h(N_2)N_{01}-r_{i1}h(N_2)N_2-(r_{i1}h(N_2)+1)N_{i1} = 0 \qquad (5\text{-}25c)$$

where $r_a = \gamma_a/\gamma_r$ and $r_{ar} = \gamma_a/\gamma_{ar}$. Coefficients r and r_{i1} are as defined earlier. These are three linear equations in N_1, N_{a1} and N_{i1}. Eliminating N_{a1} using Equation 5-25b, i.e.,

$$N_{a1} = r_{ar}N_1 \qquad (5\text{-}26)$$

the solution for N_1 is found to be

$$N_1 = \frac{rf(I_1)N_{01}}{rf(I_1)(1+r_{ar})+(1+r_a)(r_{i1}h(N_2)+1)} \qquad (5\text{-}27)$$

Similarly, the solution for the other neural population is

$$N_2 = \frac{rf(I_2)N_{02}}{rf(I_2)(1+r_{ar})+(1+r_a)(r_{i2}h(N_1)+1)} \qquad (5\text{-}28)$$

which is obtained from Equation 5-27 by reversing the subscripts. Using $h(N) = N^n$ as before, N_1 and N_2 can be calculated by solving Equations 5-27 and 5-28. Note that when $\gamma_a = 0$ and $\gamma_{ar} = 0$, this solution reduces to that found previously for the no-adaptation case (Equations 5-19).

5.5.1 Comparison with Keeler's (1968) Model

Keeler (1968) attempted to model the increase of TTS with noise exposure using a lumped-parameter circuit model. His general model consists of two exponentials:

$$TTS = TTS_{\infty}(1-ke^{-\frac{t}{\tau_1}} -(1-k)e^{-\frac{t}{\tau_2}}) \qquad (5\text{-}29)$$

The fit of this model to the Ward et al. (1958) auditory adaptation data show that for the fatigue stage, $\tau_1 = 5$ minutes and $\tau_2 = 47$ minutes. Similarly, for the recovery stage, $\tau_1 = 11.1$ minutes and $\tau_2 = 250$ minutes. Keeler's model can be derived quite simply from Equations 5-24 if (a) the inhibition terms are omitted, and (b) two adaptation levels are provided. The resulting rate equations in this case are:

$$\frac{dN}{dt} = \gamma f(I_2)N_g-\gamma_r N-\gamma_{a1}N-\gamma_{a2}N \qquad (5\text{-}30a)$$

$$\frac{dN_{a1}}{dt} = \gamma_{a1}N-\gamma_{ar1}N_{a1} \qquad (5\text{-}30b)$$

$$\frac{dN_{a2}}{dt} = \gamma_{a2}N-\gamma_{ar2}N_{a2} \qquad (5\text{-}30c)$$

The solution for N   is readily found using Laplace transform techniques (e.g., McCollum and Brown, 1965) and is of the form

$$N(t) = N_\infty + Ae^{-\frac{t}{\tau_1}} + Be^{-\frac{t}{\tau_2}} + Ce^{-\frac{t}{\tau_3}} \qquad (5\text{-}31)$$

One of the exponential terms corresponds to the excitation/de-excitation process and can be neglected when times of the order of minutes are being discussed. The remaining two decaying exponentials are identified as the two exponentials of Keeler's equation. Thus, Keeler's equation is basically a solution to a special case of the present model, and his estimates of the time constants, although derived for TTS, will be used as estimates of the time constants for selective adaptation.[3] In fact, one parameter can be eliminated from Equation 5-27 using Keeler's estimates of the time constants. Since, in Equation 5-27, $r_a = \gamma_a/\gamma_r$ and γ_r is expected to be of the order of 300 $\sec^{-1}$, r_a should be of the order of 10^{-6}. Equation 5-27 involves only the sum $1 + r_a$, so r_a can be omitted, in which case

---

[3] Since auditory adaptation/fatigue is a well known environmental hazard and a precursor to various types of hearing disorders, it is reasonable to suppose that at least some of the same physiological effects occur under repeated presentation of speech sounds. Whether or not other (e.g., phonetic) effects also exist is a moot point, and there are no data which conclusively decide either way. Since the general feeling is that adaptation is an auditory rather than phonetic effect, it seems reasonable at this point to assume that the same physiological mechanisms are involved as in TTS, and that the same time constants apply.

$$N_1 = \frac{rf(I_1)N_{01}}{rf(I_1)(1+r_{ar})+r_{i1}h(N_2)} \qquad (5\text{-}32)$$

A similar result holds for Equation 5-28.

5.5.2 Modelling the Selective Adaptation Paradigm

Equations 5-27 and 5-28 yield the values of N_1 and N_2 in response to sustained exposure to a stimulus of constant intensity. Since these are the steady state solutions, they do not contain time as an explicit parameter, and no statements can be made concerning the temporal development of (or recovery from) adaptation. Thus, to model the results of Experiments 8 and 9, it is necessary to assume that "complete" adaptation has occurred, i.e., that the boundary shifts have stabilized. Considering the number of repetitions of the adaptors and the results of auditory adaptation studies, this assumption should not be major source of error.

A further assumption which must be made is that the amount of adaptation incurred by a single presentation of a test stimulus is negligible compared to that caused by the presentation of the adaptor. Since the time constants of recovery from adaptation/fatigue are large compared to the duration of a single stimulus, after adaptation the number of cells which can be potentially excited in response to a stimulus is

$$N'_{01} = N_{01} - N_{a1} \qquad (5\text{-}33a)$$

where N_{a1} is given by Equation 5-26. (The number of excited states remaining will be assumed nil since the time constant for recovery from the excited state is expected to be very short). Using Equation 5-26, the effective number of cells available for excitation is

$$N'_{01}(I_a) = N_{01} - r_{ar} N_1(I_a) \qquad (5\text{-}34)$$

where N_1 (I_a) is the steady state level of excitation due to prolonged presentation of and adapting stimulus of intensity I_a. Thus, the neural population effectively has a reduced sensitivity as a result of the prolonged stimulation. The presentation of a test stimulus in the adapted state can therefore be modelled by computing N'_{01} from Equation 5-34 above and then using this value of N'_{01} instead of N_{01}. The boundary stimulus after adaptation is the value of α for which $N_1 = N_2$.[4]

5.5.3 Fitting the /da/-Adaptor Data

It is convenient to fit the data from Experiment 9 first since, if the model is correct, the boundary shifts for the composite adaptor can be predicted from that due to either adaptor alone. The fitting process is carried out essentially as described for the discrimination data in Section 5.4. The same method of least squares was employed: the sum of squares difference between the model and the data was calculated for each of the two subjects. To compute the predicted boundary shifts, only r_{ar} was allowed to vary. All other parameters remained set as for the AX discrimination data in Section 5.4. The adapted boundaries were obtained by computing N'_{01} from Equation 5-34 for each /daɛ/ adapator of intensity α_a , and the the solution $N_1 = N_2$ was then iteratively determined.

The results of the least squares fit are shown in Fig. 5.12. Because of the variablity of the boundary shifts for

---

[4] This assumes that there is no bias. However, any bias due to repeated presentation of an anchor stimulus will result in a shift in the same direction, and thus in this model will be indistinguishable. From the results of Simon and Studdert-Kennedy (1978) it would appear that anchor effects are generally smaller than effects due to adaptation. For this reason, and lacking a component of the model which dictates excactly how the boundary would shift due to bias, the bias is taken to be zero.

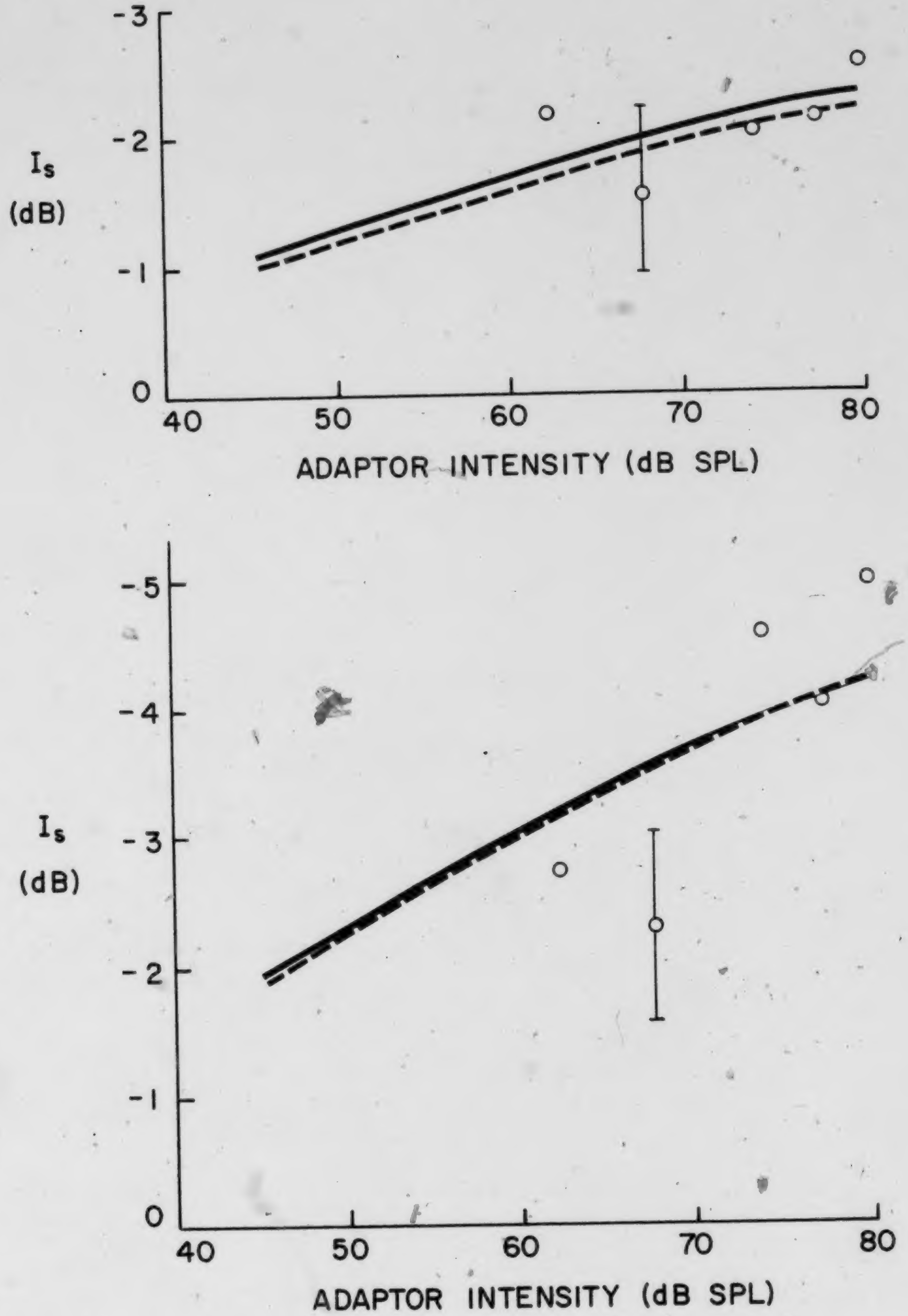


Fig. 5.12. Best-fit adaptation curves for the case of a /dae/ adaptor. The solid line represents the n=1 model and the dashed line the n=2 model. (a) subject DS (b) subject JH Vertical bar is an estimate of the limit error based on the observed reproducibility of the unadapted boundary

this experiment (which for individual subjects appears to be typical - see Ward et al., 1958), it is difficult to decide whether or not this model demonstrates the correct behaviour. Since only one parameter was allowed to vary during the fitting process, this restricts the range of behaviour of the model. However, calculations show that letting other parameters vary (e.g., r) does not substantially effect the behaviour of the model, and therefore the predicted curves of Fig. 5.12 are representative of this model. As with the AX discrimination data, part of the difficulty arises from the fact that the power law behaviour of loudness does not hold at low intensities 20 dB or so above threshold (Hellman and Hellman, 1975; Hellman and Zwislocki, 1963). In these studies, the growth of loudness increases as a power law with a larger exponent for the low intensity ranges. Hellman and Hellman suggest a modification of the power law behaviour of the form

$$f(I) = I^{\theta}(\beta - e^{-(1+\tau I)^{\frac{1}{2}} \ell n_{\beta}}) \qquad (5\text{-}35)$$

in order to account for this low intensity deviation. (Hellman and Hellman determine that 0.9 and 2.5 are representative values of β and T, respectively). This equation is based on Zwislocki's (1973) model of the firing rates of sensory neurons, and modifies the form of the loudness function for intensities within about 20 dB of threshold. This modification also has the desirable property that it changes the slope of the function f(I) from $-\infty$ to 0 for $I = 0$.

There is yet another reason for suspecting the low intensity behaviour of f(I). It was pointed out above that due to the rapid loss of signal fidelity for low intensity signals (due to fewer bits available for the digitial represenation of the signal), below approximately 20 dB or so the signal behaves as if more noise were being added to the signal. In terms of the present experimental paradigm, then, the effective threshold is perhaps only 30 - 40 dB down from the maximum signal intensity. For these reasons, it is likely that the low intensity behaviour of the composite /bae/-/dae/ signals is not adequately characterized by $f(I) = I^{\theta}$. The calculated boundary shifts shown in Fig. 5.12 are consequently too large for low intensity adaptors. To test this, a modification to f(I) was attempted:

$$f(I) = I^{\theta}(1-e^{-\beta I^{\frac{1}{2}}}) \qquad (5\text{-}36)$$

which is a simplified implementation of the Hellman and Hellman (1975) correction (Equation 5-36). Letting β be a free parameter, the adaptation model was re-computed, and the results are shown in Fig. 5.13. Some improvement in the

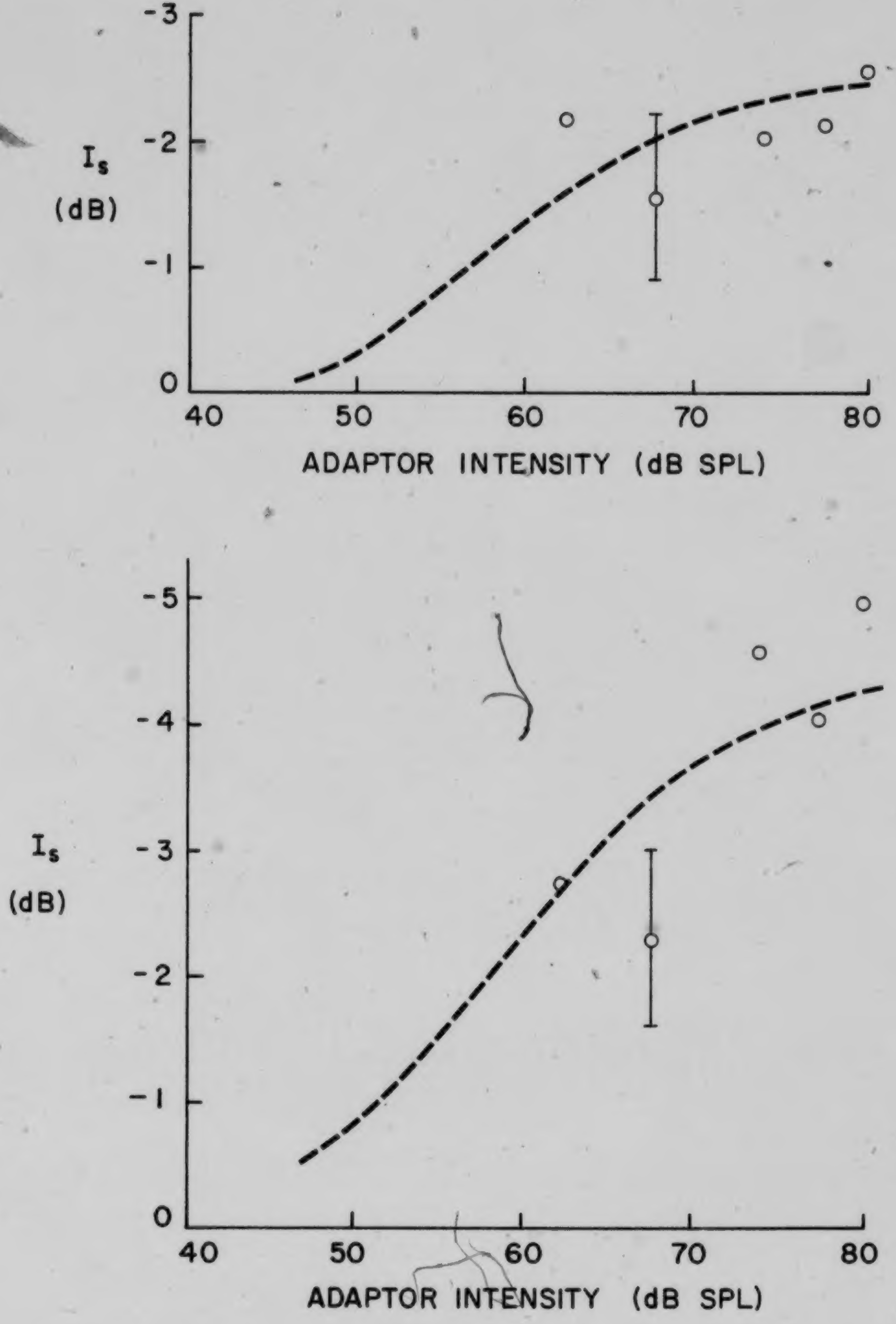


Fig. 5.13. Best-fit adaptation curves for the case of a /dae/ adaptor when the low-intensity correction for N(α) is included (calculated for n=2 model only). (a) subject DS (b) subject JH

fit of the model is observed, inasmuch as it is possible to tell with this data. Effectively, significant boundary shifts do not occur until the adaptor has an intensity of 60 dB SPL or so. This is roughly in accordance with measurements of TTS, which show that significant TTS is not produced until the adaptor intensity is somewhere in the range 60 - 80 dB (Selters, 1964; Ward et al., 1958).

The values of r_{ar} extracted during the fitting process were 0.13 for subject DS and 0.23 for subject JH. These values appear low by at least an order of magnitude, since Keeler's (1968) estimated time constants show that r_{ar} should be approximately 2. It is not clear why this is so. The fact that only one "adaptation level" was included in the model is not likely to account for such a large deviation from the expected result. However, since r controls the overall level of excitation in the neural population, it likewise controls the number of adapted cells. If r is increased, N is increased. Thus, for a constant N_a, if r is increased, then r_i must decrease. With this model, similar dispersion can be produced for various combinations of parameters, and it is difficult to anchor any one of them absolutely. So, for the present purposes it is sufficient to demonstrate that the model is capable of producing the right behaviour; more sophisticated experiments will be required to elicit more accurate values of the parameters. One such experiment would involve the change in the category boundary as a function of time, since this would allow recovery of the time constants.

5.5.4 Fitting the /bae/-/dae/ Adaptor Data

Inasmuch as the estimates of fit of the /dae/ adaptor model leaves much to be desired, approximately the right behaviour is produced. The important question now is whether or not the boundary shifts for the composite /bae/-/dae/ adaptor (Experiment 8) can be predicted on the basis of either adaptor alone. The model for the composite adaptor is virtually identical to that for the single adaptor above. The only difference is that for each adaptor α_a , both N'_{10} and N'_{20} are computed from Equation 5-35 (using α_a and $1-\alpha_a$). The predicted boundary shifts are then calculated by iterating a value of α for which $N_1 = N_2$. The calculated boundary shifts for the composite adaptor case are shown in Fig. 5.14.

Again, the model performs better for subject JH (whose boundary shifts were averaged with those for GR) than for subject DS. The most important feature of the calculated curve for n=2 is that it shows the desired inflection. The n=1 model, on the other hand, shows incorrect curvature, and therefore is eliminated from further consideration. (Calculations show that this curvature persists for the n=1 model for any range of parameters). The fact that the n=2

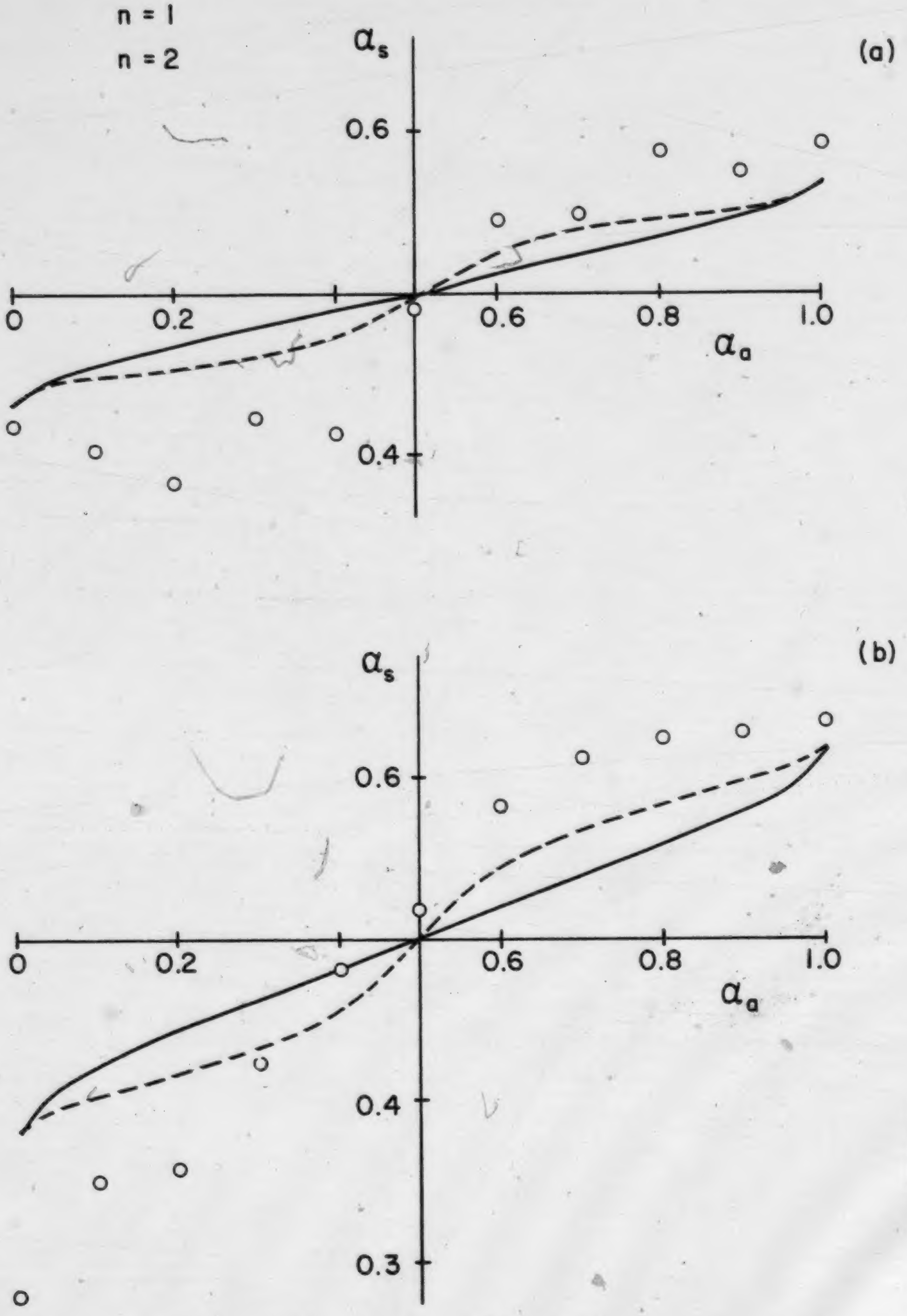


Fig. 5.14. Calculated boundary shifts for the case of a composite /bae/-/dae/ adaptor for the n=1 (solid lines) and n=2 (dashed lines) model. (a) subject DS (b) subject JH

model underpredicts the boundary shifts is not particularly disturbing in the light of Fig. 5.12. The failure of the model to properly account for the low intensity behaviour propagates into the composite-adaptor model as a reduction in the boundary shift for α close to 0 or 1 (i.e., pure /dae/ or pure /bae/ adaptors). The discrepancy is also enhanced by the fact that for subject DS, the boundary shift for $\alpha_a = 0$ is less than that for $\alpha_a = 0.1$ or $\alpha_a = 0.2$. For subject JH, the /dae/-adaptor fit (Fig. 5.12) does not produce a shift for $\alpha_a = 0$ as great as the approximate asymptotic values of Fig. 5.13, which prevents the /bae/-/dae/ boundary shift curve from attaining the large shifts necessary to improve the overall fit. Improvement in the fit of the model can be achieved by increasing r_{ar}.

The slight upturn of the predicted boundary shift curve for α_a close to 1 (and downturn close to $\alpha_a = 0$) again results from the fact that the power law $f(I) = I^\theta$ has an infinite slope at $I = 0$, which again suggests that f(I) is primarily at fault. Using f(I) defined by Equation 5-37 above, the boundary shifts were re-calculated, and the predicted boundary shifts are shown in Fig. 5.15. The predicted shifts are somewhat closer to the measured results, which means that the modification required to improve the fit of the /dae/-adaptor shifts at the same time improves the fit of the composite-adaptor data. Within the context of this model, then, the view is supported that the

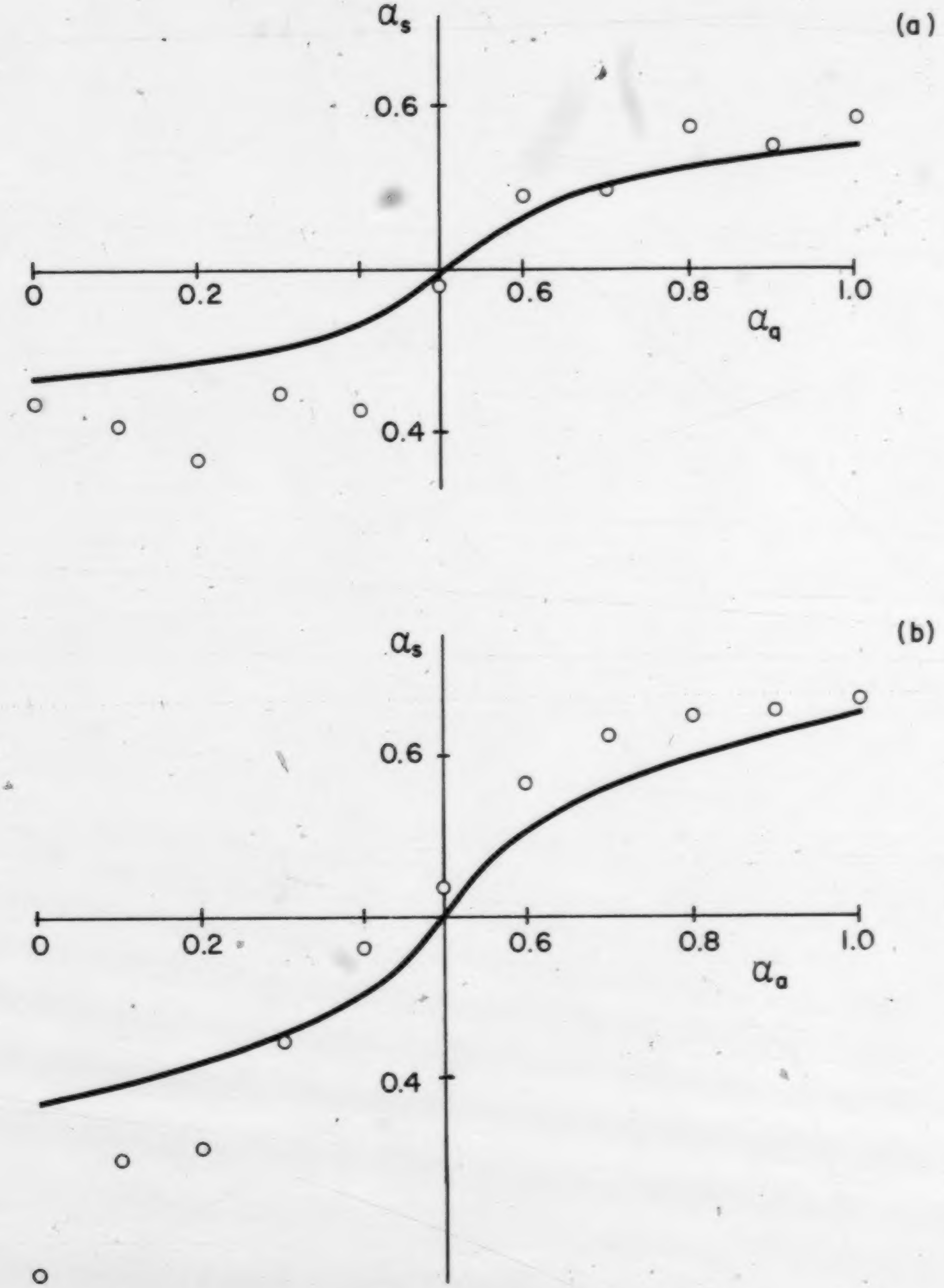


Fig. 5.15. Calculated boundary shifts for the case of a composite /bae/-/dae/ adaptor when the low-intensity correction for N(α) is included. (a) subject DS (b) subject JH

effects of the composite adaptor are basically due to that of /bae/ and /dae/ adaptors acting independently, and the curvature of the boundary shift curve (e.g., Fig. 5.15) results from the effect of the inhibition. In one sense, the effect of the inhibition is to make the system operate in a more "phonetic-like" manner.

5.6 SUMMARY

Both the discrimination model and the selective adaptation model were computed allowing only one or two parameters to vary. This places severe demands on the model, and the fits can certainly be improved by allowing more parameters to vary. This would not accomplish much, however, since the status of many of the assumptions made in the formulation of the model are unclear. In the present instance, the intent is not to derive accurate estimates of the parameters since the value of the parameter can only be as secure as the assumption which brings it into the model. What is of more interest is whether or not the model behaves in the right way. Since the simple model assuming strict power law dependence on intensity (Equations 5-2) cannot account for the observed data, the task is then to find the "just necessary additional conditions" to give the model the correct form.

The model has been cast in terms of hypothetical neural populations in order to use linear systems analysis. At best, it is a functional model, and represents "creative neural modelling". Even so, the model has a certain explanatory value. While it cannot be unambiguously claimed that this model "proves" that the /b/ and /d/ processing is carried out via essentially separate channels of analysis, it certainly appears as if this is the case. Mutual inhibition is perhaps only one of many "just necessary additional conditions", but it is a plausible inclusion to the model. It provides the suppression of the weaker signal component which seems to be indicated in Experiments 1 through 8, and provides the dispersion of the α- continuum which is necessary to account for the discrimination results. However, it will take more experiments than these to establish whether or not this is a reasonable analysis.

The present model assumes that the adaptation effects occur as a result of desensitization of a specialized neural population. Consequently, it is consistent with a "detector theory" model of selective adaptation, but it does so indirectly. Since the model assumes that the /b/ and /d/ components of the composite signal are functionally orthogonal, whatever neural entities which are responsible for the recognition of /b/ and /d/ can be treated as static templates. Thus, although the hypothetical neural populations assumed in the model are "detectors", they are so in only a limited sense. There is nothing in the model which claims that such physically distinct neural ensembles exist, only that within the present specialized circumstance they respond as if they were distinct. It is a large step from these limited detectors to detectors which span some continuum such as frequency of F_2 etc.

5.7 Extension to Dichotic Listening

Since discrimination and selective adaptation paradigms are two of the major experimental methodologies of speech perception studies, it is worthwhile to consider whether or not the present monaural fusion paradigm can aid in the interpretation of a third major source of speech perception data, binaural fusion. As will be seen, this result places further demands on the model, and although the model is observed to be inadequate to account for the complexity of this data, it certainly forms a strong basis for the interpretation of the results. Its principal virtue in the dichotic paradigm is to provide a basic framework to investigating the additional complexity required to account for central integration of monaural acoustic cues.

CHAPTER 6

BINAURAL FUSION

Depending on the type of stimuli which are presented dichotically, the resulting percept may be either fused or unfused. (The stimuli are said to be "fused" when only a single entity is perceived and "unfused" when two separate entities are perceived). Fusion of CV syllables generally occurs if the stimuli are temporally aligned and have the same fundamental frequency (Repp, 1976). The identity of the fused percept usually corresponds to one of the two stimuli which are presented, but is sometimes a phonetic mutant thereof (Cutting, 1976). Most experiments have involved dichotic contrasts of stimuli which differ in their spectral structure (see Cutting, 1976, for a taxonomy of the paradigms), with the two stimuli being presented at equal intensities. Berlin, Lowe-Bell, Cullen, Thompson and Stafford (1972), however, using dichotically presented nonsense syllables at various interaural intensities, show that a right ear superiority can exist even when the right ear signal is attentuated 15 dB or more below the left ear signal level. This indicates that ear dominance, for certain pairs of stimuli, may be relfected as a differential sensitivity to monaural inputs at some central site. Repp (1976) suggests that the degree of central interaction should be sensitive to the relative interaural intensities of the acoustic cues and also depends on the perceptual distance of the stimuli from their "prototype" values. In the monaural fusion paradigm under investigation, stimulus composition is controlled by the relative intensity of the the two signal components, and by using simultaneous monaural/binaural fusion, it should by possible to investigate in more direct fashion the role of relative interaural intensities on the central integration of speech cues.

6.1 EXPERIMENT 10: DICHOTIC PRESENTATION

The dichotic experiment described in this chapter consisted of carrying out Experiment 1 simultaneously in both ears, but reversed for the left ear. That is, as varied from 0 to 1 in the right ear, it simultaneously varied from 1 to 0 in the left ear. Thus, as the /b/ component increased in the right ear it simultaneously decreased in the left. The converse was true for the /d/ component. Fig. 6.1 illustrates the presentation of the stimuli.

This design was intended to test the sensitivity of binaural recognition to subtle differences in interaural /b/-/d/ ratios by making the combined signal (i.e., if the left and right ear intensities were added) contain an approximately equal binaural intensity of /b/ and /d/ for all values of α . Expressed in equational form, the stimuli presented to the right and left ears were

$$S_R = \alpha s_b + (1-\alpha) s_d \qquad (6\text{-}1a)$$

$$S_L = (1-\alpha) s_b + \alpha s_d \qquad (6\text{-}1b)$$

where s_b and s_d represent the time waveforms of the /b/ and /d/ formant transitions.

As an additional control parameter, the interaural intensity difference, Δ I, was also varied. For a given value of Δ I (in dB), the right and left ear stimulus combinations were scaled by factors w_R and w_L which were calculated from Δ I according to

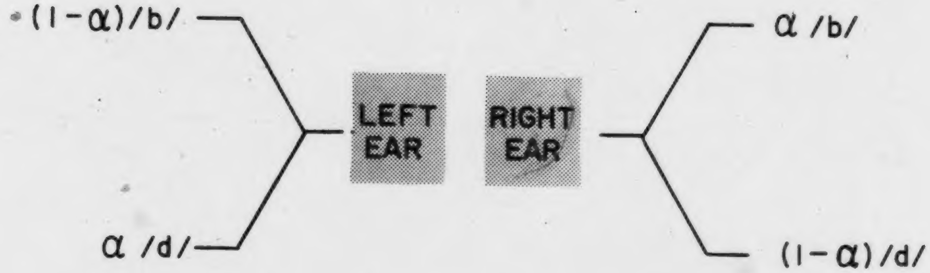


Fig. 6.1. Schematic arrangement for dichotic presentation of composite /bae/-/dae/ stimuli

$$w_L^2 = (1 + 10^{-\frac{\Delta I}{10}})^{-1} \qquad (6\text{-}2a)$$

and

$$w_L^2 = 1 - w_R^2 \qquad (6\text{-}2b)$$

(i.e., $\Delta I = 10\log_{10}(w_R^2/w_L^2)$. The signal combinations for the right and left ears were then scaled by w_R and w_L to yield

$$S'_R = w_R S_R \qquad (6\text{-}3a)$$

$$S'_L = w_L S_L \qquad (6\text{-}3b)$$

The scaling factors w_R and w_L calculated in this fashion maintained approximately equal binaural loudness independently of the value of ΔI. The twenty-one intensity-adjusted stimuli used in the ABX and AX discrimination studies were also used in this experiment to equalize the overall intensity of the composite stimuli.

Each run was conducted using a particular value of ΔI and consisted of ten presentations each of the twenty-one dichotic /bae/-/dae/ combinations. ΔI ranged from -25 dB to +25 dB in 5 dB steps, including the left monaural ($\Delta I = -\infty$) and right monaural ($\Delta I = +\infty$) cases. All aspects of the playback circuitry which might affect the right and left ear intensity differences were controlled. Matched amplifiers, filters and headphones were used to ensure equal fidelity of both playback channels and to minimize the amount of crosstalk.[1] The presentation level (80 dB SPL) was checked prior to each session and the headphones were checked for balance. The session-to-session variability was less than ±0.3 dB. Only one subject participated at a time, and wore a matched set of TDH-49 headphones (Fig. 4.8), always in the same orientation. All aspects of interstimulus timing, response collection and tabulation were as previously described for Experiment 1. The subject responded to each stimulus presentation as either /bae/ or /dae/ (or /bae/-like vs. /dae/-like).

The five subjects of Experiment 1 participated; all were right handed. Three runs for each value of I were obtained from subjects GR, JH and DS, and two from subjects GM and PA. Four or five runs were generally carried out on any particular session. The first run of any session was either left or right monaural (i.e., $\Delta I = -\infty$ or $+\infty$) and served as a practice run for the session.

The perceived stimuli were always fused, and some of the stimuli sounded like a clear /bae/ or /dae/. Most, however, had a rather elusive identity. There was the strong sensation of being able to perceive both /bae/ and /dae/ simultaneously, with the /b/ occasionally appearing to slightly lead the /d/. Controlling overt response bias was very difficult for all subjects, and the individual runs showed considerable variability. The subjects were frequently reminded to try as hard as possible to maintain a constant decision criterion.

---

[1] It was impossible to eliminate crosstalk entirely. The separation of filters and amplifiers had some effect, but the crosstalk at the headphones (as measured by a B&K artificial ear) was still only -26 dB.

The percentage /bae/ responses for all five subjects are shown as three-dimensional surfaces in Fig. 6.2. A few representative individual runs are shown in Fig. 6.3 and the averaged runs for subject GR are shown in Fig. 6.4. The two independent variables for the response surfaces of Fig. 6.2 are α , the fraction of /b/ in the right ear and Δ I, the interaural intensity difference. The profiles for Δ I= $-\infty$ and Δ I= $+\infty$ represent monaural identificaton runs and show similar, but reversed, indentification curves.

The monaural identification runs were fitted to normal ogives as previously described for Experiment 1 in order to obtain estimates of the category boundaries, I_{50}. A three-way analysis of variance was performed by concatenating the left and right ear monaural runs of Experiment 1 to those of Experiment 10. The monaural identification runs of Experiment 1 differed from those of Experiment 10 only in that the opposite earphone was open-circuited to prevent any

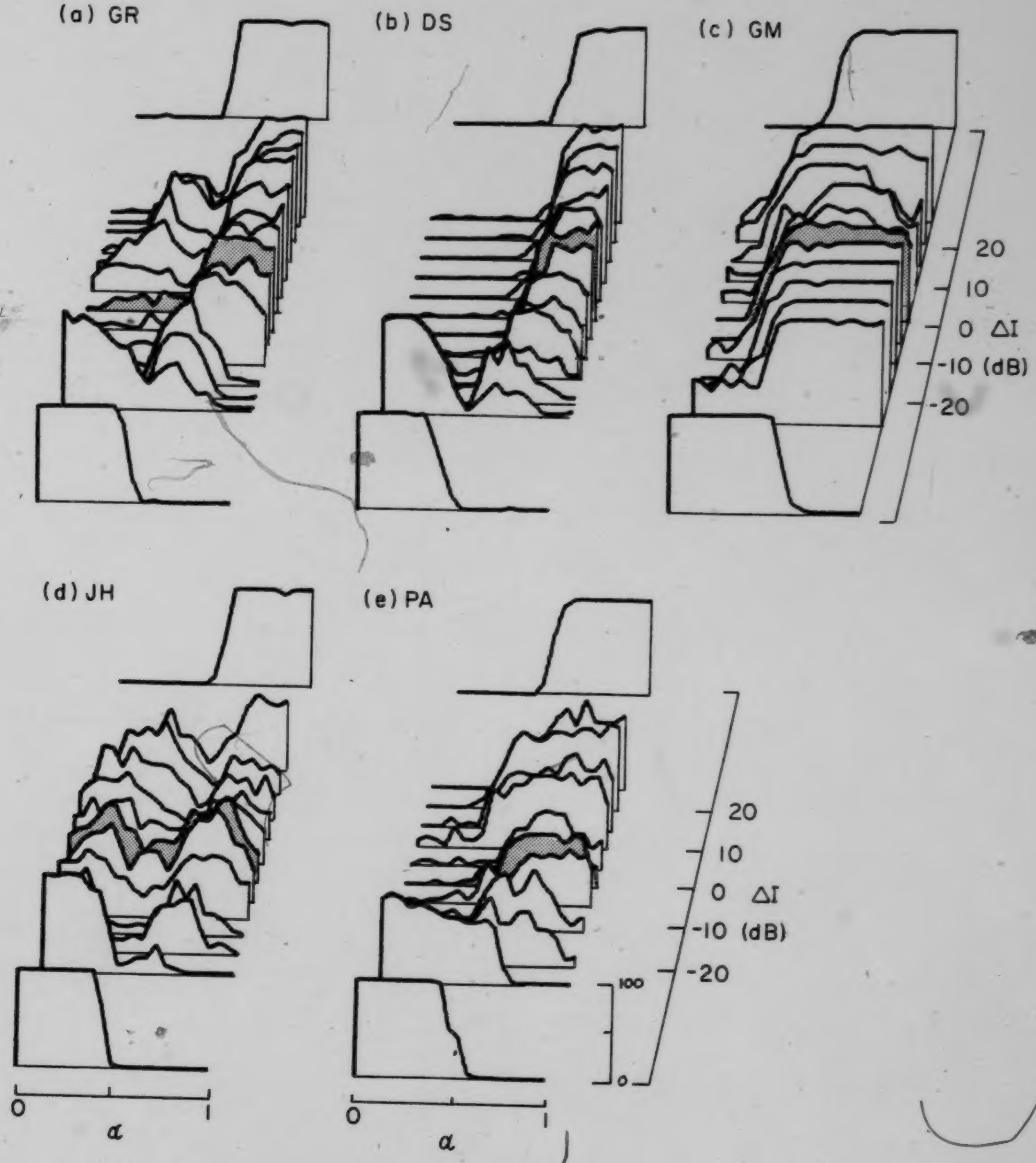


Fig. 6.2. /bae/ identification curves as a function of interaural intensity, ΔI, for the five subjects. "RM" and "LM" mark the conditions for right monaural and left monaural presentations, respectively. The shaded profiles indicate the binaural identification curves (i.e., ΔI=0 dB)

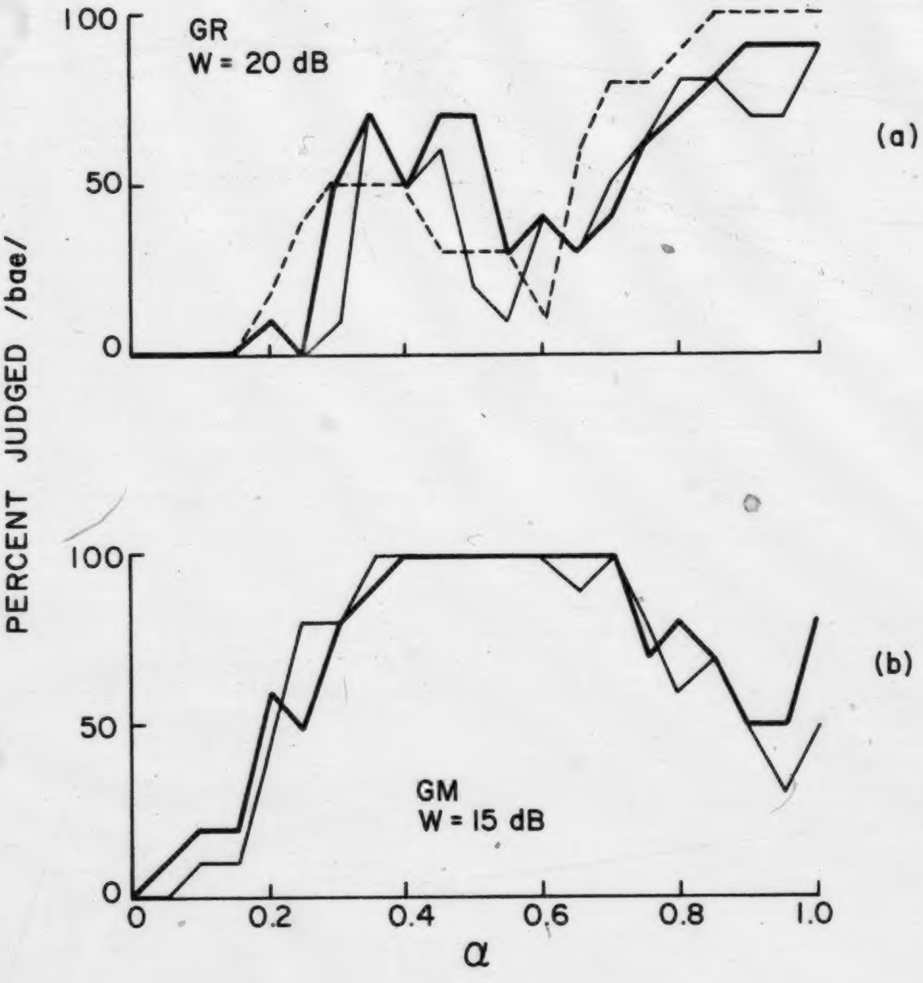


Fig. 6.3. Typical /bae/ identification curves for two subjects. (a) subject GR (b) subject GM

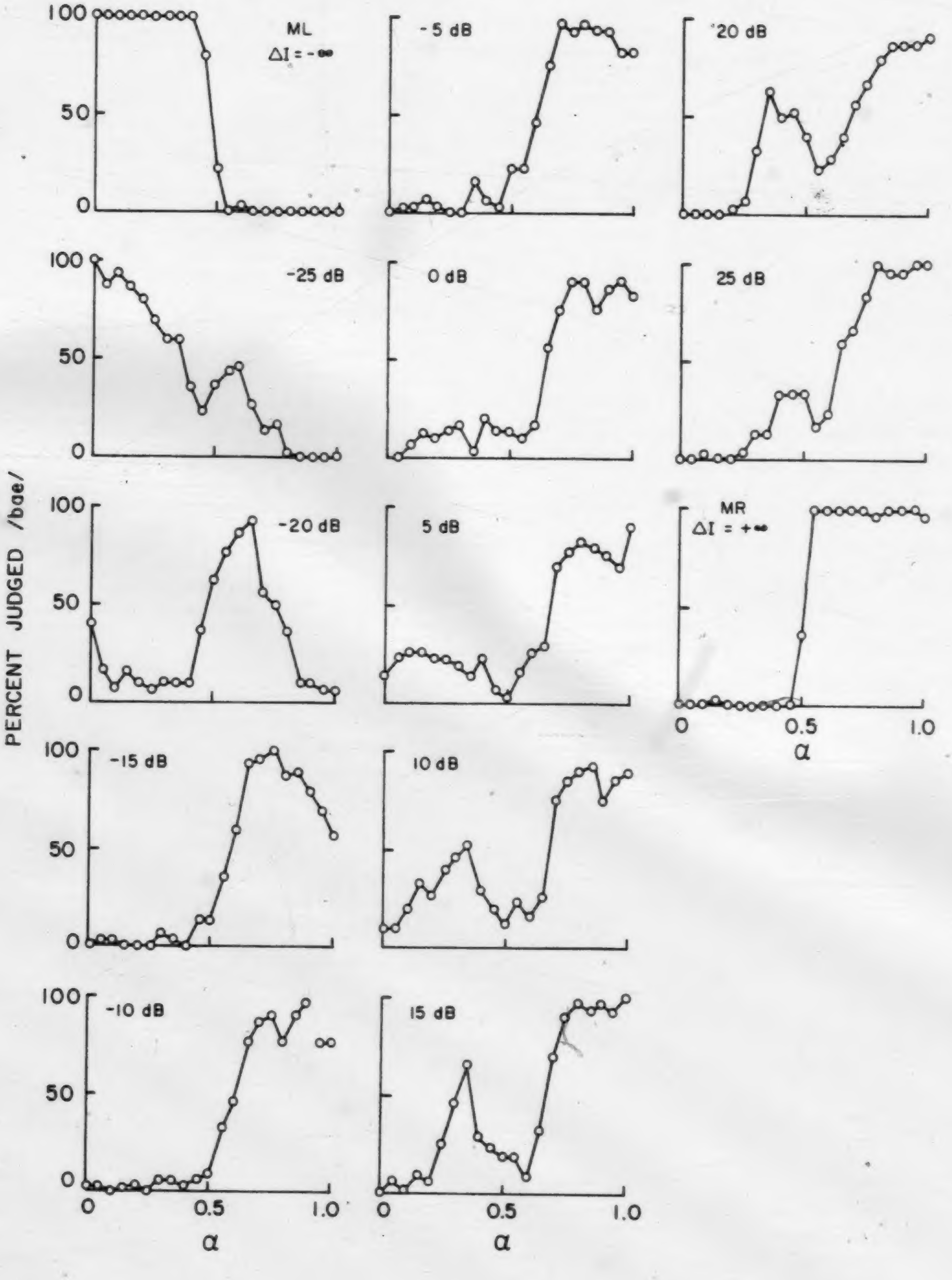


Fig. 6.4. Averaged /bae/ identification curves for subject GR for values of ΔI ranging from -∞ (left monaural) to +∞ (right monaural).

crosstalk. This condition will be referred to as M1 (one-earphone monaural) and the monaural runs of Experiment 10 will be referred to as M2 (two-earphone monaural). The predictor variables for the ANOVA were SUBJECT, EAR and CONDITION, where CONDITION was either M1 or M2. Both SUBJECT ($p<0.001$) and CONDITION ($p<0.001$) were found to be significant. EAR was not significant and there were no significant interactions. SUBJECT differences have already been discussed in Section 3.2.2 and will not be considered further. Calculation of the mean boundaries for the M1 and M2 conditions for both left and right ear cases revealed that in all but one case the M2 condition produced a value of I_{50} which was 0.5 to 1.2 dB higher than the corresponding M1 condition. The only exception occurred for subject JH for left ear M1 and M2 runs. In that case I_{50} was smaller for the M2 condition by 0.5 dB. Overall, the M2 condition was 0.82 dB higher than the M1 condition for the right ear presentations, and 0.76 dB higher for the left ear presentations. At present no explanation can be offered for this systematic shift.

6.2 EAR DOMINANCE

The profiles for $\Delta I = 0$ dB (the shaded sections of Fig. 6.2) represent the perfectly binaural case where the intensities of the composite signals delivered to the left and right ears were equal. It can be seen from Fig. 6.2

that for the perfectly binaural case the responses generally follow the stimulus in the right ear, indicating right ear dominance. The exception is subject JH who evidently demonstrates little or no ear dominance. Subjects GR, DS and GM all found the task most difficult around $\Delta I = -20$ dB (i.e., with the right ear signal 20 dB below the left ear signal). As can be seen from Fig. 6.2, this is the point where the recognition scores stopped following the right ear signal and began following the left ear signal. Thus, it appears that for these subjects the point of no ear dominance is in the vicinity of -20 dB, which is 5 - 10 dB less than that found by Berlin et al. (1972). Subject JH found all non-monaural runs to be difficult, evidently due to his lack of ear dominance.

To extract more information on possible ear dominance, the data for each subject corresponding to $\alpha = 0$ and $\alpha = 1$ were plotted as a function of ΔI (Fig. 6.5). These curves (which correspond to the sides of the 3-D surfaces in Fig. 6.2) show the percentage of responses when (a) a pure /bae/ is presented to the right ear and a /dae/ is presented to the left and (b) a pure /dae/ is presented to the right ear and a /bae/ is presented to the left.[2]

The identification curves of Fig. 6.5 show the same

---

[2] Hereafter these two conditions will be referred to as "/b/-/d/" and "/d/-/b/" respectively.

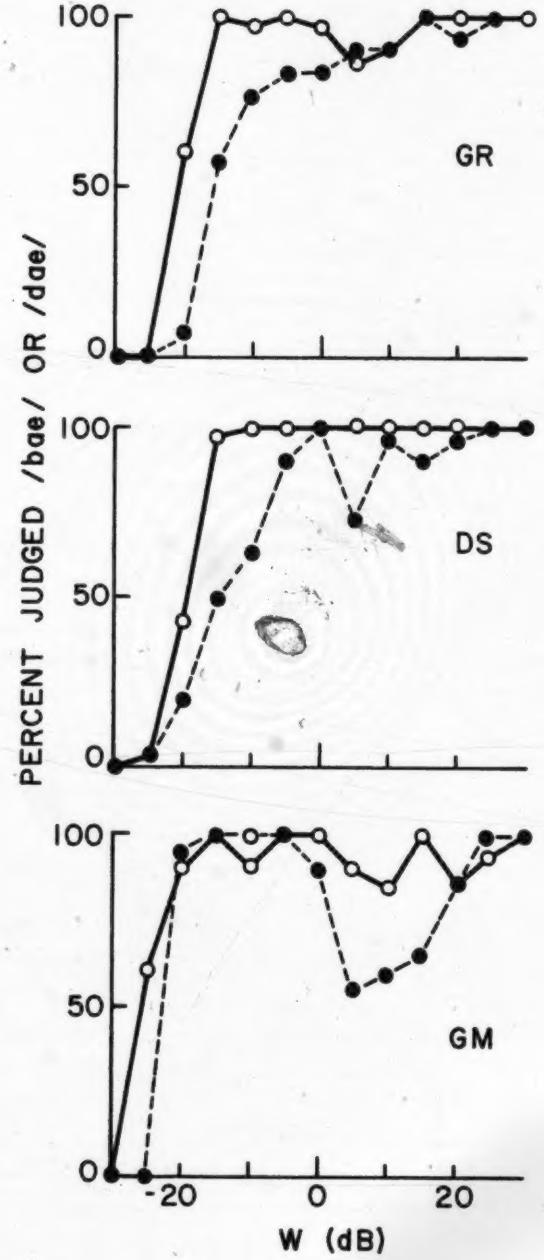


Fig. 6.5. Ear dominance curves for the /bae/-/dae/ condition (dashed lines) and /dae/-/bae/ condition (solid lines)

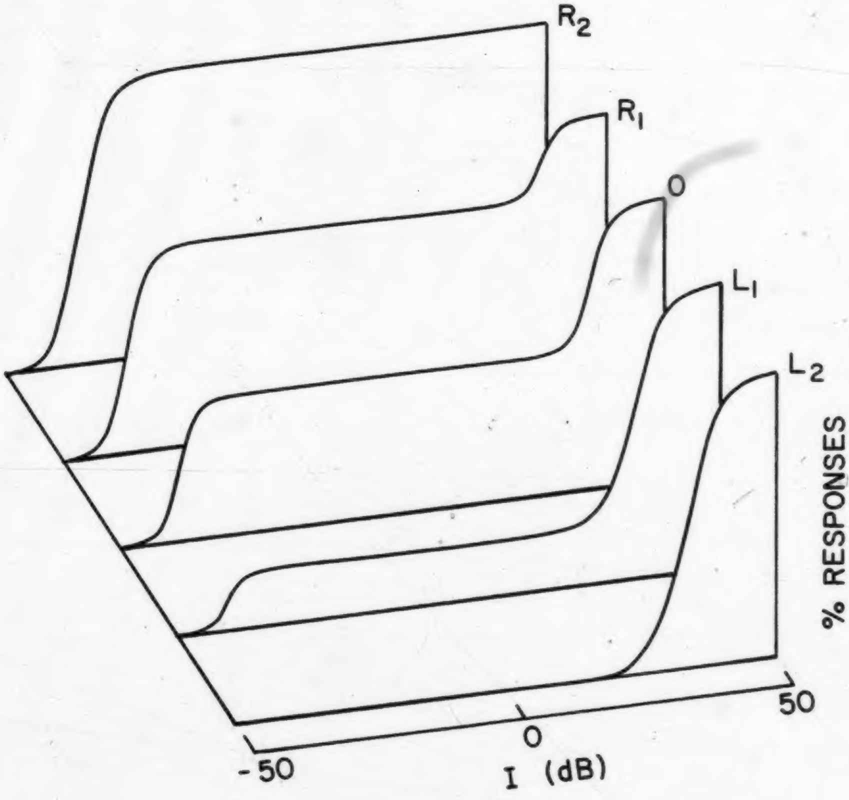


Fig. 6.6. Typical ear dominance curves for dichotic chords (after Yund and Efron, 1975)

trend as the ear dominance curves for dichotic chords obtained by Efron and Yund (1975; see also Yund and Efron, 1975). In the Efron et al. experiments, dichotic "chords" consisting of a combination of two tones of slightly differing frequencies (e.g., 1650 and 1750 Hz) are presented to the two ears. In a two interval discrimination paradigm, the lower tone is presented to the left ear simultaneously with the higher tone in the right ear, followed by the reverse configuration. The subjects' task is to identify the interval which contains the higher pitch. Efron and his co-workers have obtained ear dominance curves for these dichotic chords, and have summarized their findings as a family of ear dominance curves as shown in Fig. 6.6 (Yund and Efron, 1975). For right ear dominance, ear dominance curves such as R_1 or R_2 are obtained, where the subscript indicates increasing ear dominance. According to this classification scheme, subjects GR, DS and GM show pronounced right ear dominance. Curve "0" represents case of no ear dominance. The major feature of such curves is a broad intensity independent plateau typically extending ±30 dB. Subject JH in Fig. 6.5 evidently shows such a plateau, although in his case it only extends ±10 dB. A curious feature of the curves of Fig. 6.5 is the dip around $\Delta I = +10$ dB (with the possible exception of subjects JH and PA). The data of Efron, Dennis and Yund (1977) show just such a perturbation for hemispherectomized subjects (as well as a significant number of normal subjects), also at approximately +10 dB. The origin of this dip is unclear.

Another feature of Fig. 6.5 is the the difference between the /b/-/d/ and /d/-/b/ curves: the /b/-/d/ curve is systematically lower than the /d/-/b/ curve for all subjects except possibly GM. This difference between the two ear dominance curves represents "stimulus dominance", and can be attributed to a different ear dominance for the two stimuli (Repp, 1977). The present data indicate that, in general, there was more right ear dominance for the /dae/ than for the /bae/. Repp (1976) suggests

$$\phi_D = \frac{T_{RE(i)} - T_{RE(j)}}{T_{RE}T_{LE}} \qquad (6-4)$$

as an index of stimulus dominance, and

$$\phi_E = \frac{T_{RE(i)} - T_{LE(j)}}{T_{(i)}T_{(j)}} \qquad (6-5)$$

as an index of ear dominance (where (i) = /b/ and (j) = /d/). $T_{RE(i)}$ represents the fraction of /b/ responses when /b/ was presented in the right ear and $T_{LE(j)}$ is the fraction of /d/ responses when /d/ was presented in the left ear (see Repp, 1976, for a full definition of the variables). These indices are typically computed for the perfectly binaural case (equal intensities at the two ears). The stimulus dominance index, ϕ_D , is related to the difference in the

/b/ and /d/ scores (in the present example) for a given ear, and the ear dominance index, ϕ_E, is related to the difference in ear scores for a given stimulus. If ϕ_E is 1.0, the subject is right ear dominant, and if -1, he is left ear dominant. Presumably, when $\phi_E = 0$, the subject has no ear dominance for that particular pair of stimuli. Table 6-1 below summarizes the two indices for the five subjects, as calculated from Equations 6-4 and 6-5 for the points on the ear dominance curves corresponding to $\Delta I = 0$ dB.

TABLE 6-1

EAR AND STIMULUS DOMINANCE INDICES

SUBJECT	ϕ_E	ϕ_D
GR	0.81	-0.22
JH	-0.19	-0.37
DS	1.00	0.00
GM	0.95	-0.23
PA	0.33	-0.82

Table 6-1 shows pronounced right ear dominance for subjects GR, DS and GM and a slight left ear dominance for subject JH. These results compare favourably with a visual comparison of Fig. 6.5 with the Efron et al. ear dominance curves in Fig. 6.6. The data for subject PA are equivocal, since the /d/-/b/ curve for this subject (Fig. 6.5) shows strong right ear dominance while the /b/-/d/ curve shows slight left ear dominance. Table 6-1 shows a slight right ear dominance for this subject, but this merely reflects the averaging of the information from the two ear dominance curves. Table 6-1 also shows that in general the /d/-/b/ configuration is dominant over the /b/-/d/ condition (column ϕ_D). This also is consistent with Fig. 6.4, which shows that the /b/-/d/ curve is in general lower than the /d/-/b/ curve.

6.3 A MODEL OF BINAURAL INTERACTION

The monaural model (Section 5.2 etc.) can be extended to accommodate the binaural case following the suggestions of Repp (1976). Repp proposes a "multicategorical" model in which stimulus processing occurs in three stages: (a) auditory processing (b) multicategorical processing and (c) a higher level phonetic decision. The first stage, auditory processing, is identified in the present model as the initial transduction of stimulus energy and is represented by the excitation function f(I) in Equation 5-10. The second stage, "multicategorical processing", is the conversion into the excitation levels N_1 and N_2 of the two neural populations. The third stage is the decision process involving N_1 and N_2 and is represented by the $N_1 - N_2$

decision plane. The Yund and Efron (1977) model for pitch salience of dichotic chords is identical with the present model at this level of description.

The binaural model corresponding to Experiment 10 is represented by the two equations

$$N_2' = N_{2R} + \omega_2 N_{2L} \qquad (6\text{-}6a)$$

and

$$N_1' = \frac{\omega_1}{\omega_2} N_{1R} + \omega_1 N_{1L} \qquad (6\text{-}6b)$$

where N_{2R}, N_{1R}, etc. are the excitation levels of presumed peripheral (i.e., pre-fusional) /b/ and /d/ detectors, and are given by Equations 5-10. N_1' and N_2' are the corresponding excitations at the central level. Weighting factors ω_1 and ω_2 represent ear dominance for the /b/ and /d/ stimuli respectively. (The factor $\frac{\omega_1}{\omega_2}$ will be discussed shortly). If ω_1 or ω_2 are less than unity, then the subject is right ear dominant for those stimuli; if greater than unity, he is left ear dominant.

In this model of dichotic interactions, the central integrating mechanism involves a linear combination of the monaural /b/ and /d/ excitation levels. This results in two new variables, N_1' and N_2', which then are plotted in a binaural decision plane whose form is identical to that of Fig. 5.1. This binaural model supercedes the monaural model developed in Chapter 5. Note, however, that when input to either ear is zero, this model reduces to either a left or right monaural model. Since the analysis of variance above shows that there are no significant boundary differences between the two ears, Equations 6-6 must predict the same category boundary when the input to either ear is reduced to zero. Now, in the absence of bias, the binaural category boundary is defined by

$$N_1' = N_2' \qquad (6-7)$$

A monaural boundary occurs when either N_{2R} and N_{1R} are both zero, or N_{2L} and N_{1L} are both zero. In these cases, the boundaries are defined for a value of α such that

$$N_{2R} = \frac{\omega_1}{\omega_2} N_{1R} \qquad (6-8)$$

and

$$N_{2L} = \frac{\omega_1}{\omega_2} N_{1L} \qquad (6-9)$$

This equivalence of left and right category boundaries occurs because of the inclusion of the weighting factor in Equations 6-6. The significance of this property of the model is discussed in Section 6.3.2 below.

6.3.1 Calculating Ear Dominance Curves

Equations 6-6 are valid for dichotic presentation of composite /bae/-/dae/ stimuli. For the calculation of the ear dominance curves, however, no stimulus combinations are involved. Each of the two ears always contains either a pure /bae/ or pure /bae/, the only variable being interstimulus intensity difference, ΔI. For the /b/-/d/ condition, N and N become

$$N_2' = N_{2R} \qquad \text{and} \qquad N_1' = \omega_1 N_{1L} \qquad (6\text{-}10)$$

since $N_{2L} = N_{1R} = 0$. Similarly for the /d/-/b/ condition,

$$N_2' = \omega_2 N_{2L} \qquad \text{and} \qquad N_1' = \frac{\omega_1}{\omega_2} N_{1R} \qquad (6\text{-}11)$$

Assuming, as in Chapter 5, that the maximum stimulus intensity is unity (i.e., $\alpha = 1$ represents unit intensity of /b/), amplitude of the stimulus waveforms are simply given by the scale factors w_R and w_L (Equations 6-2). The central excitation levels are then (for the /b/-/d/ condition)

$$N_2' = N_2 \, (w_R{}^2) \qquad (6\text{-}12a)$$

$$N_1 = \omega_1 N_1 \ (1-w_R^2) \qquad (6\text{-}12b)$$

where N_1 and N_2 are given by Equations 5-10.

With Equations 6-12 it is now possible to compute the ear dominance curves. For the /b/-/d/ condition,

$$p(w_R) = \Phi\left(\frac{N_2(w_R^2) - \omega_1 N_1 \ (1-w_R^2)}{\sqrt{2}\ \sigma}\right) \qquad (6\text{-}13)$$

the probability of a /b/ being identified can be calculated for all values of w_R. (A similar equation exists for the /d/-/b/ condition). Fig. 6.7 shows representative ear dominance curves calculated from Equations 6-13 and 6-12 for various values of ω_1 $(= \omega_2)$. A consideration of the stimulus trajectory formed by varying w_R (or equivalently, ΔI) shows immediately why this model cannot generate identification curves of the required form. Fig. 6.8 shows two stimulus trajectories for $\omega_1 = \omega_2 = 1.0$ and $\omega_1 = \omega_2 = 0.3$. In each case, the stimulus trajectory crosses the decision line only once, and hence always generates an identification curve with a single inflection. This model is therefore inadequate to account for the ear dominance curves of Fig. 6.5.

Yund and Efron (1977) suggest that energy from the contralateral ear via bone conduction accounts for the intensity independence of the ear dominance curves. In the

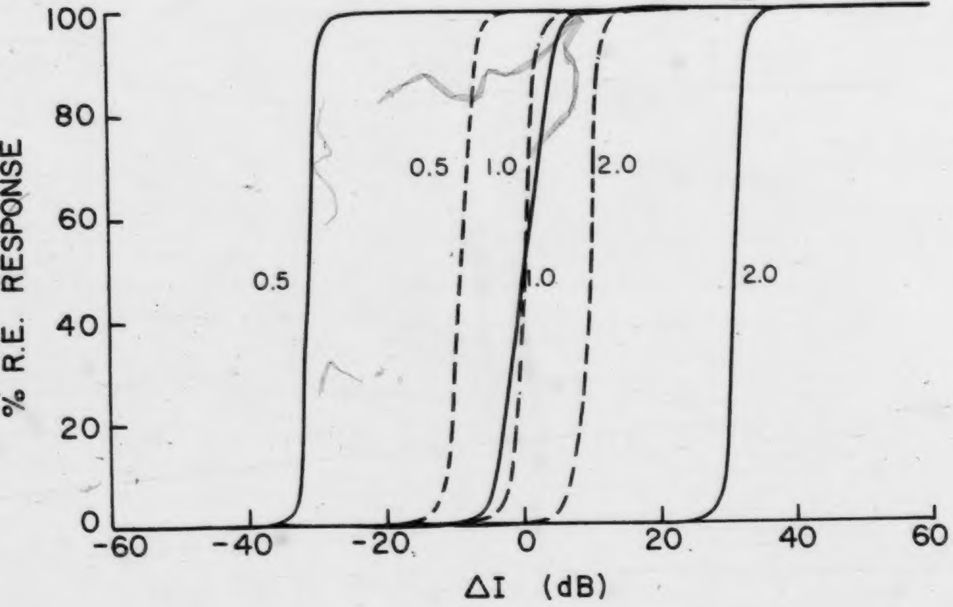


Fig. 6.7. Calculated ear dominance curves for the cases of no crosstalk (solid lines), and -26 dB of crosstalk (dashed lines). The numbers on the curves represent values of ω_1

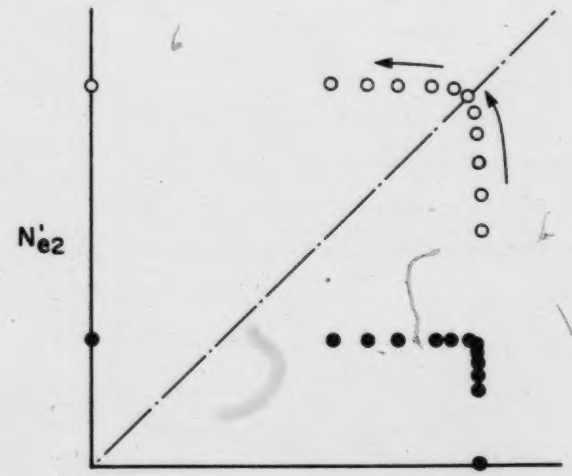


Fig. 6.8. Stimulus trajectories for $\omega_1=\omega_2=0$ (open circles) and $\omega_1=\omega_2=0.3$ (closed circles). The arrows indicate the direction of increasing ΔI

present experiment the crosstalk between playback channels was - 26 dB, which is substantially greater than the -50 dB of bone-conducted signal assumed by Yund and Efron. Crosstalk is easily incorporated into the present model by letting each monaural /b/ and /d/ detector receive C_t times the signal level of the /b/ and /d/ component of the opposing channel. For a signal level of $w_R{}^2$ in the right ear and $1 - w_R{}^2$ in the left ear, the crosstalk produces the effective intensities shown in Fig. 6.9a. Equations 6-6 then become

$$N_2' = N_2(w_R{}^2) + \omega_1 N_2(C_t w_R{}^2) \qquad (6\text{-}14a)$$

$$N_1' = \frac{\omega_1}{\omega_2} N_1(1-w_R{}^2) + \omega_1 N_2\left(C_t(1-w_R{}^2)\right) \qquad (6\text{-}14b)$$

were C_t is a factor representing the signal level from the opposite playback channel (for -26 dB, $C_t = 0.0025$). The ear dominance curves calculated assuming $C_t = 0.0025$ are shown as the dashed lines in Fig. 6.7. The primary effect of the addition of crosstalk to the model is a compression of the range of ear dominance. Therefore, crosstalk alone is insufficient to explain the intensity independence of the ear dominance curves.

There is still a major difference between the present model of binaural interaction and the model of Yund and Efron. The Yund and Efron model assumes that "... the

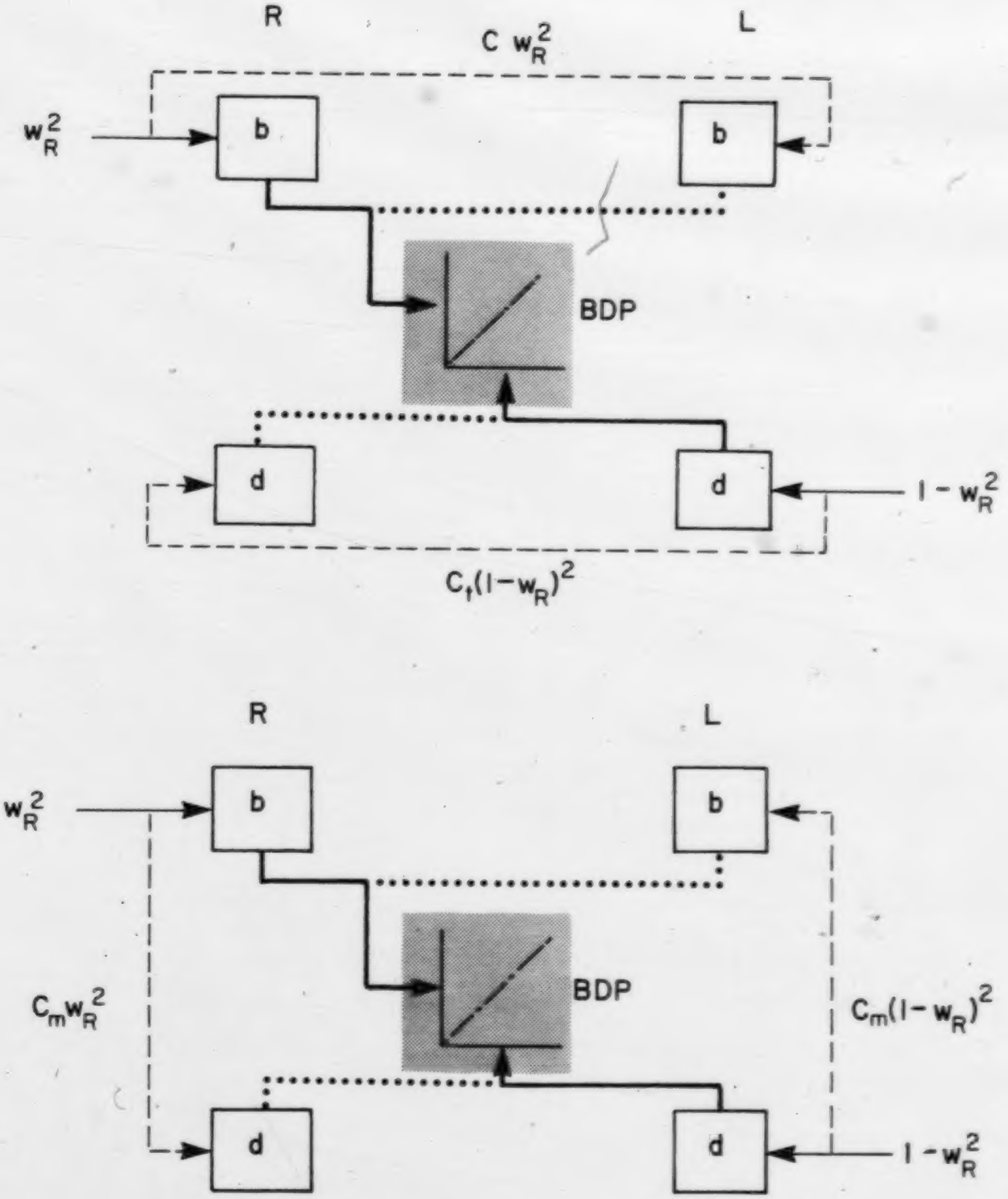


Fig. 6.9. Activation of peripheral /b/ and /d/ "detectors" (a) including crosstalk (by either bone conduction or leakage across amplifier channels), and (b) coupling of /b/ and /d/ detectors. The shaded blocks symbolize the binaural decision plane (BDP). w_R is the weighting factor which controls the interaural signal levels (Equation 6-2)

energy delivered to the ear at one frequency spreads to excite channels that are optimally sensitive to nearby frequencies." (p. 610). A similar modification to include overlap of /b/-/d/ energies can also be made to the present model.[3] The signal contributions to the peripheral /b/ and /d/ processors are shown in Fig. 6.9b. Equations 6-6 become in this case

$$N_2' = N_2(w_R^2) + \omega_2 N_2\left(C_m(1-w_R^2)\right) \tag{6-15a}$$

$$N_e' = \frac{\omega_1}{\omega_2} N_1(1-w_R^2) + \omega_1 N_1(C_m w_R^2) \tag{6-15b}$$

where C_m is a constant which determines the amount of coupling between detector inputs. When $C_m=0$, Equations 6-15 reduce to Equations 6-6. Calculated ear dominance curves using Equations 6-15 and 6-13 are shown in Fig. 6.10 for various values of ω_1 (= ω_2). These identification curves now demonstrate a functional form similar to Fig. 6.5. The reason for the sudden improvement can be understood from the nature of the stimulus trajectory. Fig. 6.11 shows that the stimulus trajectory may now cross the decision line up to three times, depending on the values of ω_1 and ω_2. It is this convoluted behaviour of the stimulus

---

[3] If /b/ and /d/ detectors span a perceptual continuum, then the assumption of coupled detectors is equivalent to assuming that one detector response function is non-zero in the region of maximum sensitivity of the other.

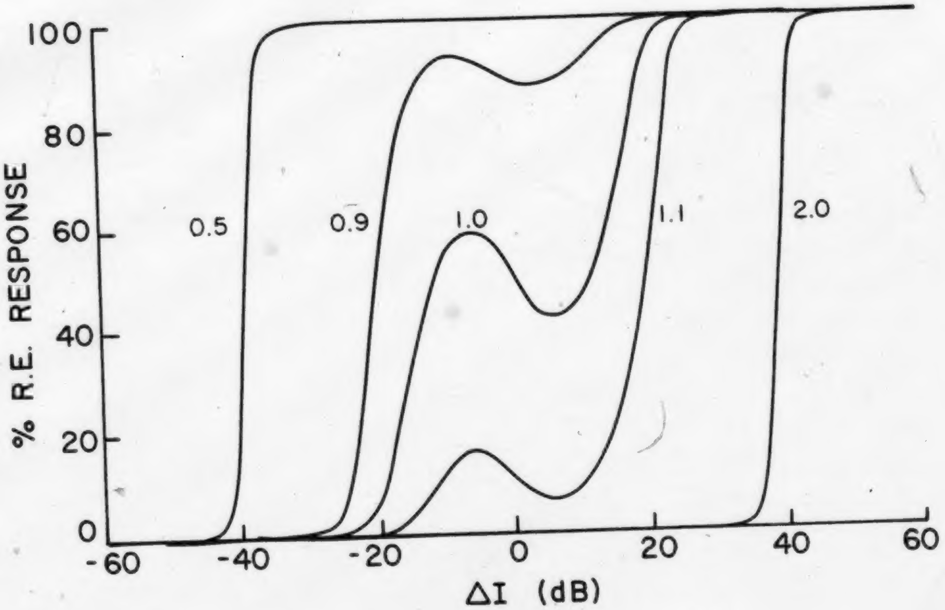


Fig. 6.10. Calculated ear dominance curves when coupling of /b/ and /d/ "detectors" is assumed

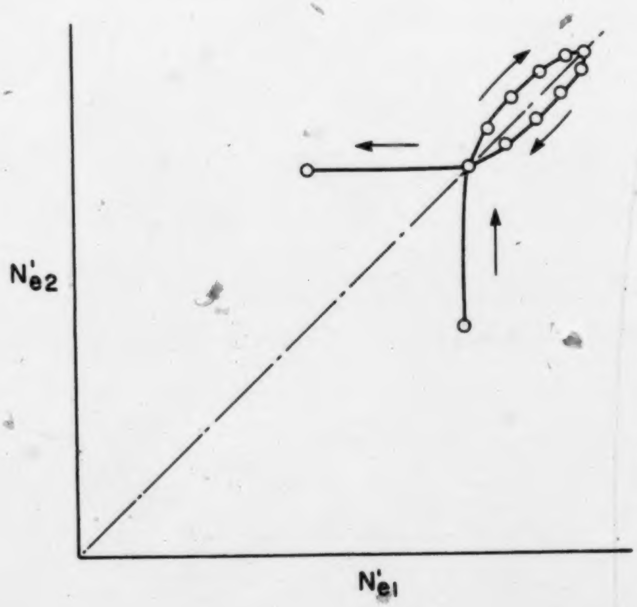


Fig. 6.11. Stimulus trajectories for $\omega_1=\omega_2=1.0$ for the case of coupled detectors. The arrows indicate the direction of increasing ΔI

trajectory which is responsible for the formation of the inflection points observed in the ear dominance curves.

These two modifications - mutual detector coupling and crosstalk between playback channels[4] - now generate a family of curves which have the required general form. To fit this model to the data in Fig. 6.5, σ , the standard deviation of the noise distribution in the binaural decision plane (e.g., Equation 6-13) was set to a nominal values of 0.01 and 0.02. ($\sigma = 0.02$ provides a better fit for the model since increasing σ smooths the ear dominance curves. $\sigma = 0.01$, as will be seen in Section 6.3.2, is a more appropriate value, and the model is calculated for both values for purposes of comparison). C_m is unknown, so it was allowed to vary. However, it was pointed out in Section 3.1 that the correlation between the /b/ and /d/ formant transitions was r= 0.32, so C_m should be of this order of magnitude. Likely it will be less since the correlation between the waveforms becomes greater as the formant transitions asymptote to the steady state values of the vowel.

An adaptive least squares fit of the model defined by Equations 6-15 and 6-13 was carried out on the ear dominance data shown in Fig. 6.5. The fitted curves are shown in Fig. 6.12 for $\sigma = 0.02$ (solid line) and $\sigma = 0.01$ (dashed line). The extracted values of ω_1 and ω_2 for the five subjects are presented in Table 6-2. The value of C_m obtained from the fitting process was 0.06, which is reasonable in view of the physical correlation between the /bae/ and /dae/ waveforms.

---

[4] Crosstalk was included in the model anyway since it was physically present during the experiment. (C_t was left set to 0.0025). Addition of crosstalk produces only a minor change in the behaviour of the model, essentially changing the value of C_m for which a given family of curves occur.

TABLE 6-2

EAR DOMINANCE WEIGHTING FACTORS

SUBJECT	($\sigma = 0.01$)		($\sigma = 0.02$)	
	ω_1	ω_2	ω_1	ω_2
GF	0.516	0.496	0.514	0.502
JH	1.038	1.022	1.050	1.023
DS	0.528	0.507	0.527	0.516
GM	0.504	0.505	0.508	0.498
PA	0.538	0.503	0.540	0.505

Repeated least squares fits using different starting values of ω_1 and ω_2 for the five subjects shows that these estimates are stable to within approximately ±0.02. The values of ω_1 and ω_2 are least reliable for subjects GM and PA due to the poor fit of the model for these two subjects. The location of the 50 percent point, i.e., the point at which the ear dominance curve rises from 0 to 50 percent is

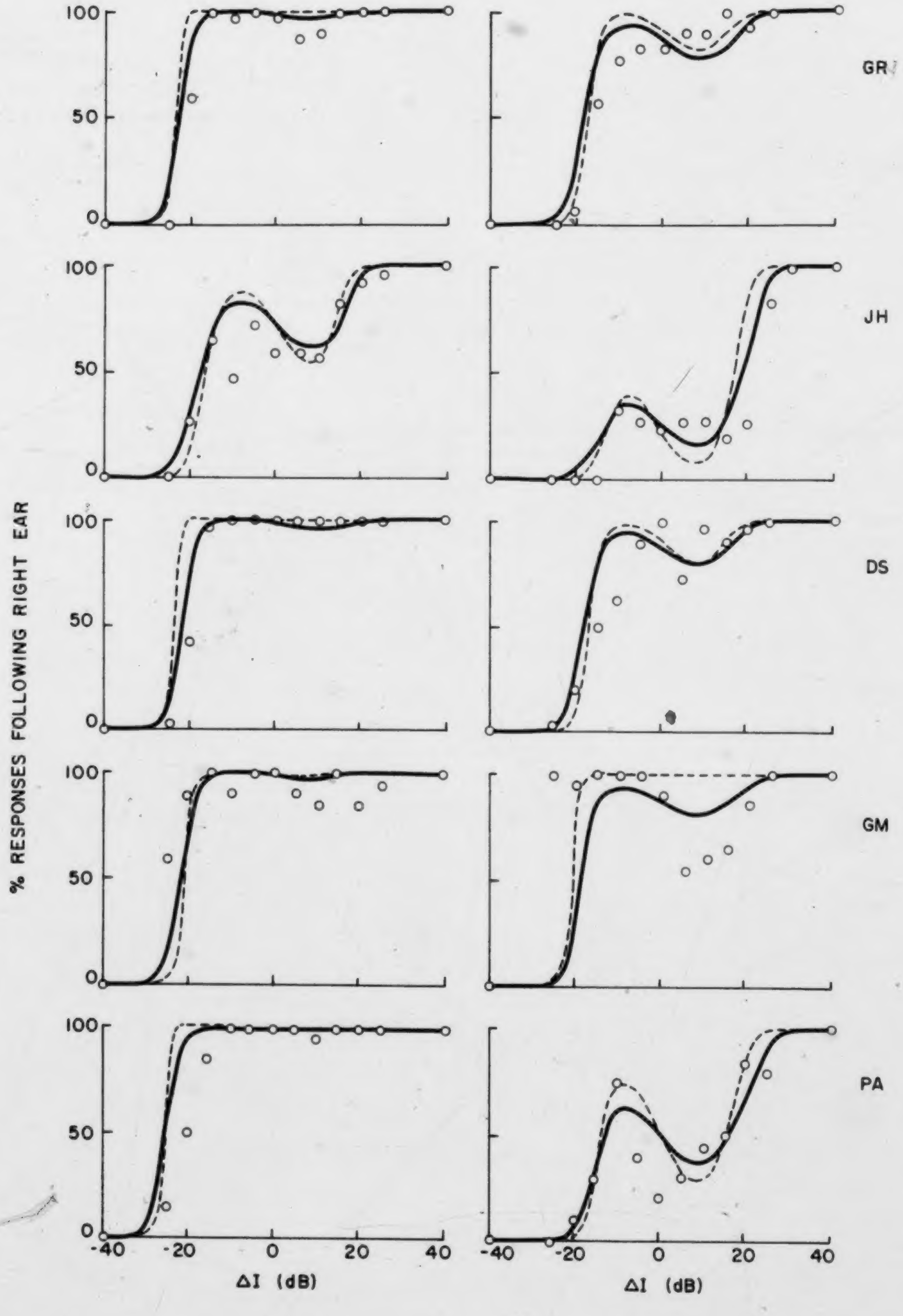


Fig. 6.12. Best-fit ear dominance curves for σ=0.02 (solid lines) and σ=0.01 (dashed lines)

the most reliable indicator of the amount of ear dominance for right ear dominant subjects, and is off by at least 5 dB for these two subjects. Evidently, subject GM should show moderate /bae/ dominance rather than the equal dominance which the model predicts. (For subjects with little ear dominance, the height of the intensity independent "plateau" is the best indicator). Fitting the model to the ear dominance curves is a stringent test of the binaural model since both the /b/-/d/ and /d/-/b/ ear dominance curves must be fitted simultaneously. A change in either ω_1 or ω_2 affects both curves, so the fit reflects an eventual compromise. This limits the ability of the model to adapt itself to the contour of the data, and consequently the fit of the model varies considerably between subjects.

6.3.2 Predicting the Category Boundaries

The binaural model (Equations 6-6) was formulated on the basis that monaural left and right ear identification runs had to predict the same category boundary. This assumption is equivalent to stating that the category boundary is determined by the relative ear dominances of /b/ and /d/, specifically the ratio $\frac{\omega_1}{\omega_2}$. This follows from the fact that a scaling of either axis of the decision plane is equivalent to changing the angle of the decision line (which is at 45 degrees for $\omega_1 = \omega_2 = 1$).[5] In the absence of response bias[6], the category boundary for right-monaural presentation is defined by the value of   for which

$$N_2' = \frac{\omega_1}{\omega_2} N_1' \qquad (6\text{-}16)$$

Using the values of ω_1 and ω_2 from Table 6-2, Equation 6-16 was solved iteratively for α, and the resulting values of I_{50} are shown below in Table 6-3 (for $\sigma = 0.01$).

TABLE 6-3

PREDICTED /tae/-/dae/ BOUNDARIES

SUBJECT	MEASURED I_{50} (dB)	PREDICTED I_{50} (dB)
GP	0.50	3.26
JH	1.57	1.23
DS	2.66	3.24
GM	-4.00	-0.03
PA	-0.52	5.61

---

[5] The angle ϕ of the decision line is given by $\tan\phi = \omega_1/\omega_2$

[6] Changing response bias is equivalent to changing the angle of the decision line. This will be indistinguishable from a change in ω_1/ω_2 and therefore can be ignored.

The predicted boundaries are of the correct order of magnitude ($I_{50} = 0$ corresponds to $\alpha_{50} = 0.5$), and the correlation between the two sets of boundaries is $r = 0.39$, which indicates that some of the boundary placement is due to the different ear dominance for the /bae/ and /dae/ stimuli. The correlation would have been higher but for the poor fits of the model for subjects GM (/b/-/d/ curve) and PA (/d/-/b/) curve.

To test out what the values of ω_1 and ω_2 should have been, the monaural identification data (the ends of the surfaces in Fig. 6.2) were fitted to the identification model given by Equation 6-16 (with w_R replaced by α). ω_1, ω_2 and were allowed to vary[7] and C_m was set to 0.06.

TABLE 6-4

EAR DOMINANCE FACTORS FROM IDENTIFICATION DATA

SUBJECT	ω_1	ω_2	σ
GR	0.480	0.477	0.007
JH	0.993	0.966	0.008
DS	0.482	0.461	0.016
GM	0.490	0.522	0.011
PA	0.477	0.480	0.141

---

[7] Actually it is not possible to recover both ω_1 and ω_2 from this fitting procedure; only the ratio ω_1/ω_2 can be found, since this alone determines the location of the boundary.

The recovered values of the three parameters are shown in Table 6-4 above (compare with Table 6-2). The average value of σ from this procedure was $\bar{\sigma} = 0.011$, which justifies the choice of $\sigma = 0.01$ used above in fitting the ear dominance data.

From the above results, the tentative conclusion can be reached that the subjects' category boundaries are determined, at least in part, by the relative amounts of ear dominance for /b/ and /d/. It is attractive to think that this is the principal reason, but a better model and more detailed data will be required before this claim can be substantiated. If it is true, then the monaural fusion paradigm affords a simple way of measuring ear dominance for a particular pair of stimuli.

6.4 PREDICTING THE DICHOTIC RESPONSES

All of the parameters necessary to describe the entire data surfaces of Fig. 6.2 have been determined in Section 6.3 above. To calculate the predicted response surfaces, Equations 6-13 and 6-6 are computed, with N_1' and N_2' defined by

$$N_2' = N_2\left(w_R{}^2(\alpha^2+C_m(1-\alpha)^2)\right) + \omega_2 N_2\left(w_L{}^2(1-\alpha)^2 + C_m\alpha^2\right) \qquad (6\text{-}17a)$$

$$N_1' = \frac{\omega_1}{\omega_2} N_1\left(w_R{}^2(1-\alpha)^2 + C_m\alpha^2\right) + \omega_1 N_{e1}\left(w_L{}^2\alpha^2+C_m(1-\alpha)^2\right) \qquad (6\text{-}17b)$$

which are the generalizations of the coupled detector equations 6-15 for the case of arbitrary /bae/-/dae/ stimulus combinations. (These equations may be easily formulated from a consideration of Fig. 6.9b). (Equations 6-15 apply only for the special cases of $\alpha = 0$ or $\alpha = 1$). Using the values of ω_1 and ω_2 given in Table 6-3 and $\sigma = 0.01$, the calculated response surfaces appear as shown in Fig. 6.13. It is observed that for the right ear dominant subjects (GF, DS and GM) the model has basically the required form. The model produces a poor representation of the data of subject JH, however. Part of this deviant behaviour can be attributed to the inability of the model to fit the ear dominance curves discussed previously. However, the most serious defect of the model is its failure to predict the two "ridges" observed in the JH data, as well as the failure to predict the "bump" for subject GF. The model performs moderately well for subject DS, but this is strictly a result of the simple data structure for that subject. The calculated response surfaces for subjects GM and PA are far from their actual data surfaces, primarily due to the incorrect values of ω_1 and ω_2 found previously. The effect of the ratio ω_1/ω_2 can be seen in Fig. 6.13 as a displacement of the inflection point for the left and right monaural identification runs.

The model captures the coarse features of the data of Fig. 6.2 and fits the edges of the data surfaces (i.e., the

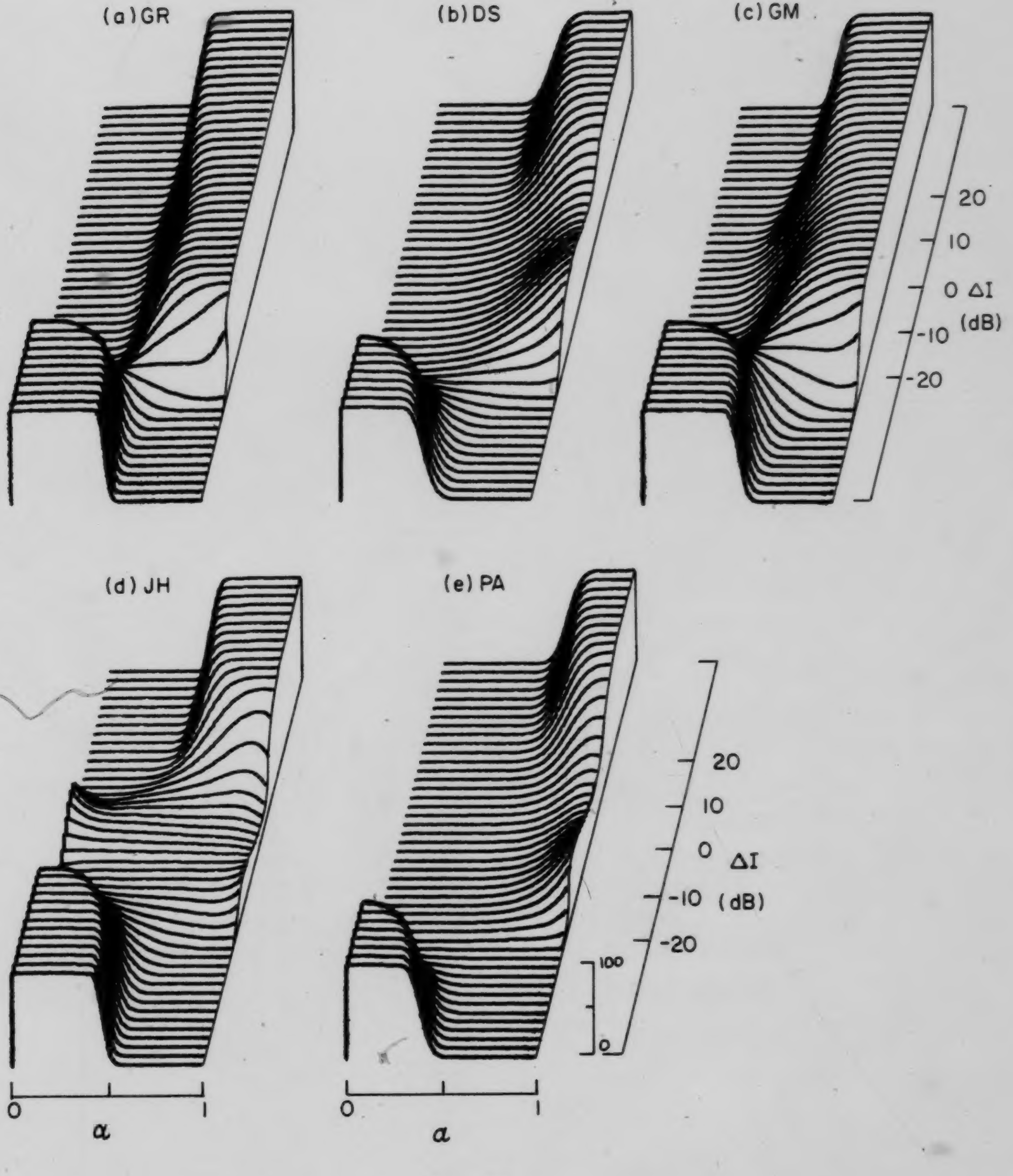


Fig. 6.13. Calculated /bae/ identification curves for the five subjects (compare with Fig. 6.2). "RM" and "LM" mark the conditions for right monaural and left monaural presentation, respectively

two ends and the two sides) fairly well. The point where the responses stop following one ear and start following the other is reasonably well predicted in all cases. Interestingly, the slight shift in the transition region between approximately $\Delta I = -5$ and $+10$ dB for subjects GR, DS and PA is also generated by the model. A slight bulge in the surface occurs for subject GM, approximately where her data also demonstrate a slight deviation. These are minor effects of the model, but major considerations for a model which truly accounts for all dichotic interactions.

In summary, the model performs fairly well for pronounced right ear dominant (and presumably pronounced left ear dominant) subjects, and not at all well for subjects with little or no ear dominance. Clearly the no ear-dominance case is the most stringent test of this (or any other) model, and until the model can adequately account for the effects observed in the no ear dominance case, the model cannot be considered successful. Accounting for all of the subject variations observed in the data of Experiment 10 will be a challenge for any model.

6.5 SUMMARY

The binaural model developed in this chapter follows the Yund and Efron (1977) model for pitch salience of dichotic chords and shares many of the same features.

Repp's (1976) suggestions for "multicategorical processing" in which peripheral excitations combine before a central phonetic decision is made appears to be basically substantiated. His suggestion that a given stimulus partially excites neighbouring detectors was found to be a key ingredient in the model. Even though the model is not entirely successful, it provides a basic measure of understanding of why the data have the form they do. The failures of the model themselves provide insight into what information must be obtained to disambiguate between possible models of dichotic interactions.

It is not exactly clear why the model does not perform better. Lack of dependence of σ on interaural intensity difference is one possibility since the ear dominance curves produce a better fits if σ is allowed to become larger. However, increasing σ simultaneously destroys the fit for the rest of the model, and these contradictory requirements of large and small σ indicate a serious shortcoming of the model. One possible solution to the problem which has not yet been investigated is to include inhibition, as was done in Chapter 5. Enhancing the dispersion along the α - continuum will allow σ to become larger and yet maintain the same discriminability (as shown in Section 5.4). Whether or not this will also generate the internal structure of the dichotic reponse surfaces is unknown at present.

CHAPTER 7

SUMMARY AND CONCLUSIONS

7.1 SUMMARY

In this thesis, several basic speech perception phenomena have been investigated in an attempt to bring them together under a single conceptual rubric. A new experimental probe - monaural fusion - has been developed for this purpose which has several advantages over existing experimental techniques. First, it uses real speech, and thus avoids the contentious issue of whether or not subjects can readily identify isolated stimuli as speech tokens (e.g., Barclay, 1972). Because natural speech tokens are used, recognition of the endpoint stimuli on the continuum is perfect for all subjects, and considerably aids the interpretation of the results. Second, the monaural fusion paradigm manipulates a well-defined physical variable: relative intensity. This permits an exact specification of the independent variable - a variable often not identifiable in speech perception studies when VOT or F_1-F_2 continua are involved. ("Stimulus number" does not constitute a well-defined independent variable).

The four experimental paradigms investigated in this thesis show one general conclusion: the /bae/ and /dae/ components of a mixed /bae/-/dae/ simulus are treated as if they were delivered over essentially separate auditory channels. This supports the view that /b/ and /d/ are recognized by functionally separate neural entities, and this independence of processors is responsible for the fact that this continuum is categorically perceived. Whether or not these neural entities represent "detectors" which also span the F_1-F_2 continuum cannot be decided with this data, but the data are certainly consistent with this view. The results of Experiment 10 - by virtue of the model of dichotic interactions - suggests that the two neural populations are not quite separate. In the detector view of /b/ and /d/ perception, this is equivalent to stating that one detector response function has non-zero sensitivity in the region of maximum sensitivity of the other (e.g., Fig.

2.12). This is consistent with the general view that in order to account for the boundary shifts under selective adaptation, the detector response functions must overlap.

The experimental investigations of Chapters 3 and 4 and the subsequent modelling of Chapter 5 show that stimulus combinations of the type described produce perceptual effects which do not require the existence of hypothetical phonetic levels of processing for their explanation. The point of the model is not to deny the existence of phonetic levels of processing, but merely to attempt to account for these basic phenomenon in terms of more basic psychophysical processes. While it may be attractive to invoke phonetic processing as an explanation for any phenomenon involving speech or speech-like signals, this should be done only when it can be successfully demonstrated that lower level processing is insufficient to predict the observed effects.

The model of the perception of mixed /b/-/d/ stimuli should be interpreted in this light. The use of the term "detector" in this case is limited to mean a functionally distinct neural ensemble which characterizes the nervous system's response to the entire /b/ or /d/ acoustic pattern. This form of detector does not necessarily correspond to a "feature detector" per se, but may perhaps represent the combined effects of complexes of feature detectors. It is not perfectly clear what level of processing these hypothetical neural populations are supposed to represent. While they may represent dedicated neural ensembles which monitor auditory input for certain select acoustic configurations, this is not the only possible interpretation. If every acoustic signal results in a neural representation which is unique (e.g., a kind of "neural spectrogram"), then this neural population may be part of the internal signal representation itself. The functional independence of the two neural populations then directly follows from the degree of orthogonality of the stimuli: two signals which do not have appreciable spectral overlap (e.g., separated by one or more critical bands) will likewise not have appreciable communality of neural exciations. The model developed in this thesis is invariant to this assumption. In either interpretation, for the purposes of modelling it is necessary to assume that some global statistic (e.g., total excitation) is the variable which participates in higher level decisions. This, of course, is more of a mathematical requirement for the purposes of modelling rather than an imposition on the nature of signal processing by the nervous system. Whatever the nature of this global statistic, it represents a condensation of information, and thus can be used to represent the output of a "detector". Since it is a continuous variable, it follows from this model that the output of a detector is not a binary value, but rather has a value which may be increased or decreased in either of two ways: (1) the intensity of the stimulus may be increased or decreased, or (2) the spectro-temporal composition of the stimulus may be altered towards or away from the "prototype" value. This implies a certain equivalence of the monaural fusion paradigm of this thesis and the classic F_1-F_2 paradigm used in synthetic speech studies, and is one possible explanation for the concordance of the effects shown for discrimination, selective adaptation and binaural fusion in this thesis and the same effects observed with F_1-F_2 continua. If this is the case, then the model developed in Chapters 5 and 6 is just a special case of a more general model which incorporates a spectro-temporal continuum as well.

The model was developed incrementally in Chapters 5 and 6 in an attempt to justify the inclusion of certain components of the model. The essential features of the model will be reviewed here. The basic model is that of a simple two-detector configuration as shown in Fig. 7.1. It was shown that this model did not demonstrate the required dispersion unless a mutual inhibitory component was added. The form of this inhibition is quite arbitrary, and the fact that it produced the required dispersion cannot be considered proof of its existence since a suitable modification of the general intensity response of the detectors themselves (function f(I) in Equation 5-10) can

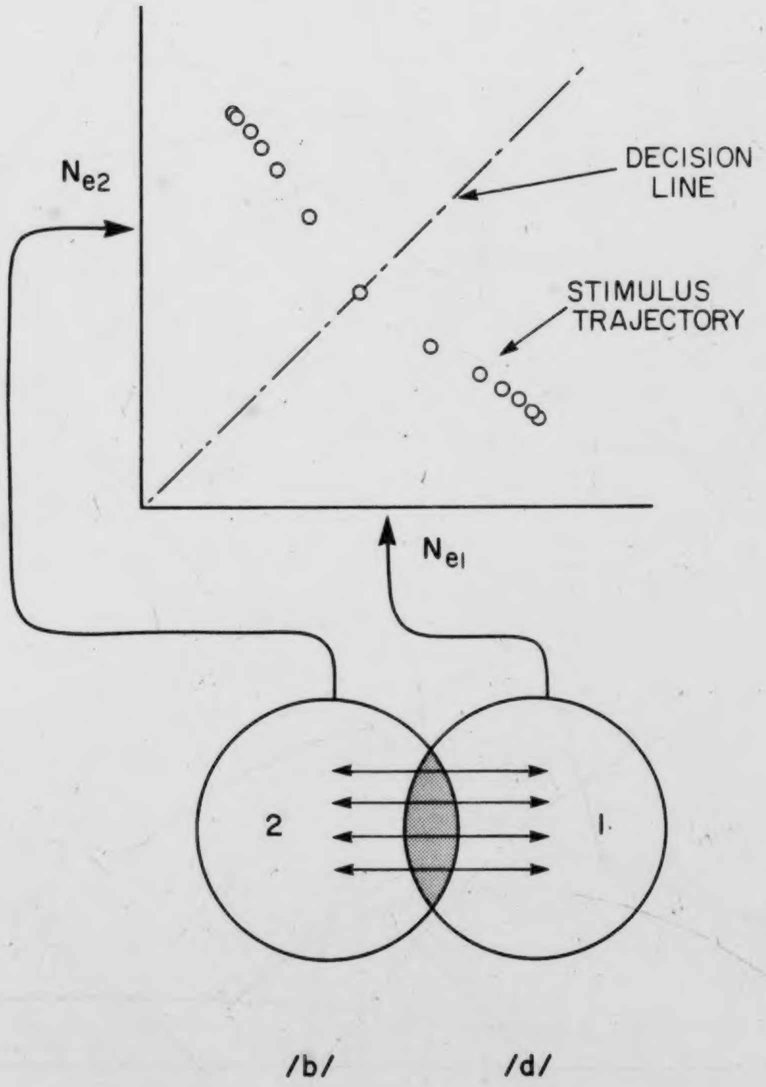


Fig. 7.1. Schematic processing of /bae/-/dae/ composite signals. The shaded area respresents possible coupling of the /b/ and /d/ processors. Arrows represent possible mutual inhibition of processor outputs

accomplish the same end. Thus, while there is evidence that some form of interaction exists between the /b/ and /d/ processors, it is not necessarily inhibition. The important point to note is that the requirement of enhanced dispersion to account for the discrimination results simultaneously imparts the correct curvature to the boundary shift curve of the selective adaptation study (Section 5.5). Furthermore, it was also shown that a modification of the low intensity behaviour of f(I) to eliminate the implausible large dispersion at low intensities was also sufficient to further increase the curvature of the boundary shift curve. This merely accentuates the fact that the behaviour of the model is crucially dependent on the intensity response function (f(I) in Equation 5-10). Until the intensity response function is known with more certainty, it will be impossible to state exactly the nature of the interaction.

The model of selective adaptation is not unreasonable in the light of parallels with auditory adaptation and fatigue. The inclusion of a special "adaptation level" in the state diagram of Fig. 5.11 is only one way of modelling the decrease in sensitivity of a neural population. This is likely to be only one component of the total effects of adaptation and fatigue. Other effects such as hair cell dysfunction or metabolic changes in the cochlea itself may account for other aspects of auditory adaptation and fatigue, but since their effects on threshold shifts or phonetic boundary shifts will be similar, it is difficult to separate the individual sources (Elliott and Fraser, 1970). A model of adaptation based on rate processes (e.g., Equations 5-24) at least provides the basic exponential behaviour typically observed in auditory adaptation studies (Keeler, 1968). The important aspect of the model is that selective adaptation be modelled in a way which produces functional dependence of boundary shift on adaptor intensity and duration (optionally number of adaptor presentations). Only the intensity dependence has been analyzed in this thesis; the temporal dependence of boundary shift is of equal importance, but will require detailed data on the recovery of phonetic boundary shifts. Such data are currently not available in the literature, but will be required for a more comprehensive understanding of selective adaptation.

The binaural extension of the model involves three basic assumptions, the most important of which is that the peripheral excitation levels are simply summed at some central site. The second assumption is that different stimuli have different ear preferences, and the third assumption is that the /b/ and /d/ components of the stimuli each partially excite their opposing detector. These three concepts are sufficient to explain the main effects observed in the data. Inasmuch as this model is only tentative, it produces the clear prediction that the placement of the subjects' category boundaries along the relative intensity (/bae/-/dae) continuum is determined solely by the relative ear dominances of /b/ and /d/. This is a prediction of the model which is partially substantiated by the present data, and it will be of great interest to conduct detailed experiments to analyze this particular aspect of dichotic interactions. In this vein, it should be noted that it is also possible to include selective adaptation into the binaural model (trivially by substituting Equations 5-33 for 5-10) and thereby predict the effects of adaptation on the ear dominance curves. This has not yet been done, but the results should help to clarify the nature of dichotic interactions as well as those of the adaptation effects themselves (see Cooper, 1974).

A major component of the model is the "decision plane", which represents a two-dimensional perceptual signal space in which the outputs of the detectors are displayed (e.g., Fig. 5.1). This is as much a conceptual aid in visualizing the behaviour of the model as it is a statement of the decision-making process. With the assumption of circular normal probability density functions (pdf's) in this plane, all integrations of the pdf's reduce to a single dimension (Abramowitz and Stegun, 1967, p. 956). Consequently, the model does not actually require postulating such a decision plane, but doing so allows the behaviour of the model in the discrimination, selective adaptation, etc. paradgims to be analyzed as geometric operations on the stimulus trajectory. This provides a measure of insight into the nature of the model which cannot be obtained by viewing the same operations in a single dimension.

In summary, this model has an explanatory value which transcends the dubious nature of some of its assumptions. Its principle virtue is that it allows four related phenomena - identification, discrimination, selective adaptation and binaural fusion to be interpreted in terms of a common set of parameters. The additional assumptions which each experimental paradigm requires provides a certain measure of insight into these phenomena. The model cannot be considered completely successful, of course, but the failures of a model are as important as its successes. The primary function of a preliminary model such as this is to formalize in mathematical terms many of the proposals which are proposed in the literature, and in so doing clarify the implications of these proposals. The advantage of having a model - any model - is that it enables future experiments to be directed at obtaining information which will resolve the issues which the model raises.

7.2 DIRECTIONS FOR FURTHER RESEARCH

The research described in this thesis describes a new experimental probe for speech perception studies, and the various experiments were conducted in an attempt to understand the nature of the probe itself. This is a mandatory requirement before the experimental paradigm can be used to probe other aspects of speech perception. The understanding of the phenomenon of monaural fusion arises from the model which was constructed to explain the experimental effects. This model possesses the required behaviour for not unreasonable values of the parameters, but suffers from the indeterminability of some of its components. Thus, future research should be primarily dedicated to resolving the various assumptions which define the model. Once a successful model of the monaural fusion paradigm is created, the model may then be generalized to other continua, and in so doing may help explain some of the diverse phenomena observed in speech perception studies. It will be especially advantageous to employ the same experimental technique using simpler stimuli, e.g., rising and falling tones, to investigate whether or not the sharp transition between categories is a result of their spectro-temporal structure or whether it is contingent on the stimuli belonging to "competing" speech categories.

Only four experimental paradigms are described in this thesis, but others have been investigated at an informal level. The possibilities of the monaural fusion paradigm as a probe are clearly not exhausted. For instance, it is possible to construct simultaneous signal mixtures in more than one position in a stimulus. Two stimuli, /baeb/ and /daed/, can be mixed to produce /baeb/, /baed/, /daeb/ and /daed/ and simultaneous boundaries for the syllable-initial and syllable-final consonant mixtures can be obtained. A recognition experiment using such stimuli will yield information on how the initial consonants affect perception of the final consonants, and vice versa. Furthermore, a selective adaptation study using such stimuli will produce information on the effects of syllable-initial adaptation on syllable-final consonant perception in a manner much more direct than that of, e.g., Ades (1974).

Another possible experiment is to set the stimulus components to the boundary values for a subject, and then to selectively increase the intensities on various portions of the speech waveform. The perceptual salience of that acoustic information can then be determined by observing the effect on the location of the category boundary. Preliminary experiments along this line have been conducted. Varying the relative intensities of each of the pitch periods of the /b/ and /d/ waveforms shows that the effect on the boundary is greatest for the first pitch period, and decreases approximately exponentially for later pitch periods, a result which would be expected if the rate of change of formant frequency was an important acoustic cue for /bae/ and /dae/.

Higher level judgments can be investigated using the monaural fusion probe. The experiments of Cutting (1975; Cutting and Day, 1975) concerning the effects of semantic influences on the boundary placements can be replicated using consonant mixtures rather than dichotic presentation of entire words. It will be interesting to observe whether or not the semantic plausibility of a phrase in any way influences the placement of the category boundary. If it does, this may provide evidence of anticipatory "tuning" of the speech-decoding system.[1]

Lastly, it should be noted that the "Necker-cube"-like phenomena created by monaural fusion is a phenomenon of interest for its own sake. Informal up-down tracking experiments show a hysteresis zone several dB in width, which is consistent with the limits of the /bae/-/dae/ boundary shown in Fig. 3.11. The forced choice identification curve may then only be the result of averaging these two separate identification curves, as shown in Fig. 7.2. It is possible to conduct this same experiment in the visual domain using a Necker cube or other reversible figure, and similar hysteresis effects should be obtained

---

[1] Monaural fusion does not occur for all possible consonant mixtures. When /ba/ is mixed with /la/, for example, the endpoint stimuli are still /b/ and /l/, but intermediate stimuli tend to have the qualities of both. Stimuli for which the components are approximately equal sometimes have the identity /bla/, which is similar to the findings of Cutting (1975).

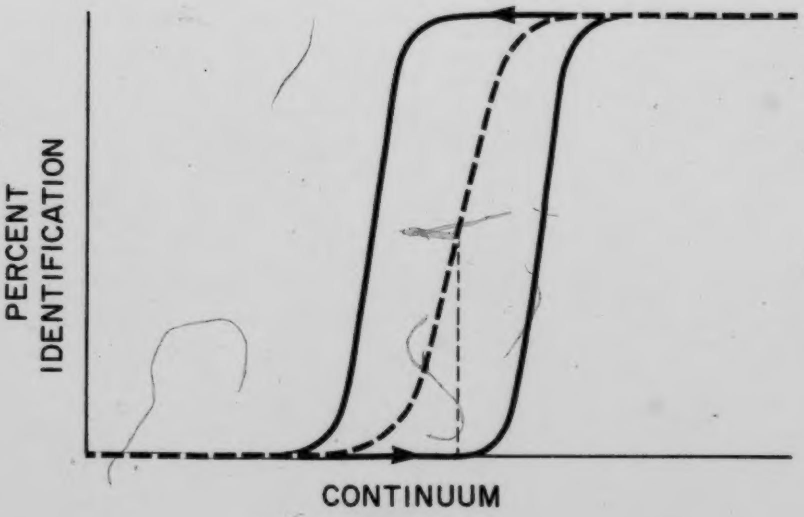


Fig. 7.2. Identification curves obtained using a sequence of stimuli either ascending or descending the continuum (solid lines marked with arrows). The dashed line represents the identification curve as would be measured by a forced-choice test drawing stimuli randomly from the continuum

(cf Taylor and Aldridge, 1974).

These are but a few of the possibilities; there are many others. The monaural fusion paradigm (with variation of relative intensities) is a very general technique which can be employed at the psychophysical level using simple tonal stimuli, or at higher levels using more complex speech stimuli. In either case, the effects are probably interpretable at a psychophysical level and thus a complete understanding of this phenomenon may help to understand the psychophysics of speech perception.

REFERENCES

Abbs, J.H and Sussman, H.M. Neurophysiological feature detectors and speech perception: A discussion of theoretical implications. J. Speech Hear. Res. ,1971, 14:23-36.

Abramowitz, M. and Stegun, I.A., EDS. Handbook of Mathematical Functions. New York, Dover, 1965.

Ades, A.E. Bilateral component in speech perception. J. Acoust. Soc. Am., 1974a, 56, 610-616.

Ades, A.E. Vowels, consonants, speech and nonspeech. Psych. Rev., 1977, 84, 524-530.

Ainsworth, W.A. Mechanisms of selective feature adaptation. Percept. and Psychophys., 1977, 21, 365-370.

Allen, J. AND Haggard, M. Perception of voicing and place features in whispered speech: a dichotic choice analysis. Percept. and Psychophys., 1977, 21, 315-322.

Alpern, M. Rhodopsin kinetics in the human eye. J. Phyisiol., Lond., 1971, 217:447-471.

Anderson, J.A., Silverstein, J.W., Ritz, S.A. and Jones, R.S. Distinctive features, categorical perception and probability learning: Some applications of a neural model. Psych. Rev., 1977, 84, 413-451.

Barclay, J.R. Noncategorical perception of a voiced stop: a replication. Percept. and Psychophys., 1972, 11, 269-273.

Berlin, C.I., Lowe-Bell, S.S., Cullen, J.K. Jr., Thompson, C.I. and Stafford, M.R. Is speech "special" Perhaps the temporal lobectomy patient can tell us. J. Acoust. Soc. Am., 1972, 52, 702-705.

Berlin, C.I. and McNeil, M.R. Dichotic Listening. In *Contemporary Issues in Experimental Phonetics.* Lass, N.J., Ed. New York, Academic Press, 1976. pp.327-387.

Blechner, M.J., Day, R.S and Cutting, J.E. Processing two dimensions of nonspeech stimuli: The auditory-phonetic disctinction reconsidered. *J. Exp. Psych*: *Human Percept. and Perform.*, 1976, *2*, 257-266.

Blumstein, S.E., Stevens, K.N. and Nigro, G.N. Property detectors for bursts and transitions in speech perception. *J. Acoust. Soc. Am.*, 1977, *61*, 1301-1313. Botsford, J.H. Theory of TTS. *J. Acoust. Soc. Am.*, 1968, *44*, 352(A).

Burns, E.S. and Ward, W.D. Categorical perception - phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical INTERVALS. *J. Acoust. Soc. Am.*, 1978, *63*, 456-468.

Carney, A.E., Widin, G. and Viemeister, N. Noncategorical perception of stop consonants Differing in VOT. *J. Acoust. Soc. Am.*, 1977, *62*, 961-970.

Cohen, J. and Little., W., IV. Neural model for the Steven's power law. *Percept. and Psychophys.*, 1971, *10*, 269-270.

Conte, S.D. *Elementary Numerical AnalysiS*. New York, McGraw-Hill, 1965.

Cooper, W.E. Adaptation of phonetic feature analyzers for place of articulation. *J. Acoust. Soc. Am.*, 1974, *56*, 617-627.

Cutting, J.E. Aspects of phonological fusion. *J. Exp. Psych.*, 1975, *104*, 105-120.

Cutting, J.E. Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psych. Rev.*, 1976, *83*, 114-140.

Cutting, J.E. and Day, R.S. The perception of stop-liquid clusters in phonological fusion. *Jour. of Phonetics* , 1975, *3*, 9-23.

Cutting, J.E. and Dorman, M.F. Discrimination of intensity differences carried on formant transitions varying in extend and duration. *Percept. and Psychophys.*, 1976 , *20*, 101-107.

Cutting, J.E. and Rosner, B.S. Categories and boundaries in speech and music. Percept. and Psychophys., 1974, 16, 564-570.

Cutting, J.E., Rosner, B.S. and Foard, C.F. Perceptual categories for music-like sounds: Implications for theories of speech perception. Quart. J. Exp. Psych., 1976, 28, 361-378.

Davis, H., Morgan, C.T., Hawkins, J.E., Jr., Galambos, R., and Smith, F.W. Temporary deafness following exposure to loud tones and noise. Acta Oto-Lar., 1950, Suppl. 88.

Dorman, M.F. Discrimination of intensity differences in formant transitions in and out of syllable context. Percept. and Psychophys., 1974, 16, 84-86.

Durlach, N.I. and Braida, L.D. Intensity perception. I. Preliminary theory of intensity resolution. J. Acoust. Soc. Am., 1969, 46, 372-383.

Efron, R. and Yund, E.W. Dichotic competition of simultaneous tone bursts of different frequency III. The effect of stimulus parameters on suppression and dominance functions. Neuropsychologia., 1975, 13, 151-161.

Efron, R., Dennis, M. and Yund, E.W. The perception of dichotic chords by hemispherectimized subjects. Brain and Lang., 1977, 4, 537-549.

Efron, R. and Yund, E.W. Ear dominance and intensity independence in the perception of dichotic chords. J. Acoust. Soc. Am., 1976, 59, 889-898.

Eimas, P.D. and Corbit, J.D. Selective adaptation of linguistic feature detectors. Cog. Psych., 1973, 4, 99-109.

Eimas, P.D., Siqueland, E.R., Juszyck, P. and Vigorito, J. Speech perception in infants. Science, 1971, 171, 303-306.

Elman, J.L. Response bias account of selective adaptation. Paper presented at the 1977 meeting of the Acoustical Society of America, Miami Beach, Florida.

Elman, J.L. Perceptual origins of the phoneme boundary effect and selective adaptation to speech: A signal detection theory analysis. J. Acoust. Soc. Am., 1979, 65, 190-207.

Foreit, K.G. Linguistic relativism and selective adaptation for speech: A comparative study of English and Thai. Percept. and Psychophys., 1977, 21, 347-351.

Freides, D. Do dichotic listening procedures measure lateralization of information processing or retrieval strategy Percept. and Psychophys., 1977, 21, 259-263.

Fry, D.B., Abramson, A.S., Eimas, P.D. and Liberman, A.M. The identification and discrimination of synthetic vowels. Lang. and Sp., 1962, 5, 171-189.

Fujisaki, H. and Kawashima, T. On the modes and mechanisms of speech perception. Ann. Rep. Eng. Res. Inst. 1969, 28, University of Tokyo, 67-73.

Fujisaki, H. and Kawashima, T. Some experiments on speech perception and a model for the perceptual mechanism. Ann. Rep. Eng. Res. Inst. 1970, 29, University of Tokyo, 207-214.

Green, D.M. and Swets, J.A. Signal Detection Theory and Psychophysics. New York, John Wiley and Sons, 1966.

Hanson, V. I. Within-category discrimination in speech perception. Percept. and Psychophys., 1977, 21, 423-430.

Hellman, W.S. and Hellman, R.P. Relation of the loudness function to the intensity characteristics of the ear. J. Acoust. Soc. Am., 1975, 57, 188-192. Hellman, R.P. and Zwislocki, J.J. Monaural loudness function at 1000 Hz and interaural summation. J. Acoust. Soc. Am., 1963, 35, 856-865.

Hillenbrand, J.M. Intensity and repetition effects on selective adaptation to speech. Research on Speech Perception, 1975, 2 (Bloomington: Indiana University), 56-137, (Cited in Simon and Studdert-Kennedy, 1978).

Hirsh, I.J. and Ward, W.D. Recovery of the auditory threshold after strong acoustic stimulation. J. Acoust. Soc. Am., 1952, 24, 131-141.

Hood, J.D. Studies in auditory fatigue and adaptation. Acta Oto-Lar., Suppl. 92.

Keeler, J.S. Compatible exposure and recovery functions for temporary threshold shifts - mechanical and electrical models. Jour. Sound Vibr., 1968, 7:220-235.

Kimura, D. Cerebral dominance and the perception of verbal stimuli. Can. Jour. Psych., 1961, 15, 166-171.

Kimura, D. Functional asymmetry of the brain in dichotic listening. Cortex, 1967, 3, 163-178.

Kuhl, P.K. and Miller, J.D. Speech perception in early infancy: discrimination of speech sound categories. J. Acoust. Soc. Am., 1975, 58, S58.

Kuhl, P.K. and Miller, J.D. Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. Science, 1975, 190, 69-72.

Liberman, A.M., Harris, K.S., Hoffman, H.S., and Griffith, B.C. The discrimination of speech sounds within and across phoneme boundaries. J. Exp. Psych., 1957, 54, 358-368.

Lipetz, I.E. The relation of physiological and psychological aspects of sensory intensity. In Handbook of Sensory Physiology, Vol. I: (Principles of Receptor Physiology), New York, Springer-Verlag, 1971, pp. 191-225.

Locke, S. and Kellar, L. Categorical perception in a nonlinguistic mode. Cortex, 1973, 9, 355-369.

Luce, R.D. Thurstone's Discriminal processes fifty years later. Psychometrika, 1977, 42, 461-489.

Macmillan, N.A., Kaplan, H.L. and Creelman, C.D. The psychophysics of categorical perception. Psych. Rev., 1977, 84, 452-471.

Marks, L.E. Visual brightness: Some applications of a model. Vis. Res., 1972, 12, 1409-1423.

Marks, L.E. Sensory Processes. New York, Academic Press, 1974.

McCollum, P.A. and Brown, B.F. Laplace Transform Tables and Theorems. New York, Holt, Rinhart and Winston, 1965.

Miller, G.A. The perception of short bursts of noise. J. Acoust. Soc. Am., 1948, 20, 160-167.

McGovern, K. and Strange, W. The perception of /r/ and /l/ in syllable-initial and syllable-final position. Percept. and Psychophys., 1977, 21, 162-170.

Miller, J., Wier, C., Pastore, R., Kelly, W. and Dooling, R. Discrimination and labeling of noise-buzz sequencies with barying noise-lead times: An example of categorical perception. J. Acoust. Soc. Am., 1976, 60 , 410-417.

Miller, J.L. Properties of feature dectectors for speech: Evidence from the effects of selective adaptation on dichotic listening. Percept. and Psychophys., 1975, 18, 389-397.

Miller, J.L. and Eimas, P.D. Studies on the selective tuning of feature detectors for speech. Jour. of Phonetics, 1976, 4, 119-127.

Miller, J.L. and Eimas, P.D. Studies on the perception of place and manner of articulation: A comparison of the labial-alveolar and nasal-stop distinctions. J. Acoust. Soc. Am., 1977, 61, 835-845.

Miller, J.L., Eimas, P.D. and Root, J. Properties of feature detectors for place of articulation. Paper presented at the meeting of the Acoustical Society of America, Spring, 1977.

Miyawaki, K., Strange, W., Verbrugge, R. Liberman, A.M., Jenkins, J.J. and Fujimura, O. An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English. Percept. and Psychophys., 1975, 18, 331-340.

Morse, P.A. The discrimination of speech and nonspeech stimuli in early infancy. Jour. Exp. Child Psychol., 1972, 14:477-492.

Pastore, R. Categorical perception: A critical reevaluation. In Hearing and Davis: Essays Honouring Hallowell Davis. Hirsh, S.K., Eldredge, D.H., Hirsh, I.J. and Silverman, S.R. (Eds.), St. Louis, Mo., Washington University, 1976.

Pastore, R.E., Ahroon, W.A., Baffuto, K.J., Friedman, C., Puleo, J.S. and Fink, E.A. Common-factor model of categorical perception. J. Exp. Psych: Human Percept. and Perform., 1977, 3, 686-696.

Petty, J.W., Fraser, W.D. and Elliott, D.N. Adaptation and loudness decrement: A reconsideration. J. Acoust.

Soc. Am., 1970, 47, 1074-1082.

Pierce, J.R. and Gilbert, E.N. On AX and ABX limens. J. Acoust. Soc. Am., 1958, 30, 593-595.

Pisoni, D.B. Auditory and phonetic memory codes in the discrimination of consonants and vowels. Percept. and Psychophys., 1973, 13, 253-260.

Pisoni, D.B. Auditory short-term memory and vowel perception. Mem. and Cog., 1975, 3, 7-18.

Pisoni, D.B. and Lazarus, J.H. Categorical and noncategorical modes of speech perception along the voicing continuum. J. Acoust. Soc. Am., 1974, 55, 328-333.

Pollack, I. and Pisoni, D. On the comparison between identification and discrimination tests in speech perception. Psychonomic Sci., 1971, 24, 299-300.

Repp, B.H. Identification of dichotic fusions. J. Acoust. Soc. Am., 1976, 60, 456-469.

Repp, B.H. Measuring laterality effects in dichotic listening. J. Acoust. Soc. Am., 1977, 62, 720-737.

Repp, B.H. Dichotic competition of speech sounds: The role of the acoustic stimulus structure. J. Exp. Psych: Human Percept. and Perform., 1977, 3, 37-50.

Repp, B.H., Liberman, A.M., Eccardt, T. and Pesetsky, D. Perceptual integration of acoustic cues for stop, fricative and affricate manner. HASKINS LABORATORIES: Stat. Rep. Speech. Res. 1978, SR-53, 61-83.

Roufs, J.A.J. Dynamic properties of vision - V: Perception lag and reaction time in realtion to flicker and flash thresholds. Vis. Res., 1974, 14, 853-869.

Sawusch, J.R. Peripheral and central processes in selective adaptation of place of articulation in stop consonants. J. Acoust. Soc. Am., 1977, 62, 738-750.

Sawusch, J.R. and Pisoni, D.B. Response organization to selective adaptation to speech sounds. Percept. and Psychophys., 1976, 20, 413-418.

Scott, A.C. Neurodynamics: A critical survey. Jour. Math. Psych., 1977, 15, | PAR.3.

Selters, W. Adaptation and fatigue. J. Acoust. Soc. Am., 1964, 36, 2202-2209.

Simon, H.J. Anchoring and selective adapation of phonetic and nonphonetic categories in speech perception. Unpublished Ph.D. dissertation. (Cited in Simon and Studdert-Kennedy, 1978).

Simon, H.J. and Studdert-Kennedy, M. Selective anchoring and adaptation of phonetic and nonphonetic continua. J. Acoust. Soc. Am., 1978, 64, 1338-1357.

Small, A.M. Auditory adaptation. In Modern Developments in Audiology. Jerger, J., Ed., New York:, Academic Press, 1963.

Small, A.M., Brandt, J.F. and Cox, P.G. Loudness as a function of signal duration. J. Acoust. Soc. Am., 1962, 34, 513(L).

Sorkin, R.D. Extension of the theory of signal detectability to matching procedures in psychoacoustics. J. Acoust. Soc. Am., 1962, 34, 1745-1751.

Sorkin, R.D. and Pohlmann, L.S. Some models of observer behaviour in two-channel auditory signal detection. Percept. and Psychophys., 1973, 14, 101-109.

Stevens, S.S. Neural events and the psychophysical law. Science, 1970, 170, 1043-1050.

Stevens, S.S. Sensory power functions and neural events. In Handbook of Sensory Physiology, Vol. 1, Lowenstein, W.R., Ed., New York, Springer-Verlag, 1971.

Stevens, S.S. and Hall, J.W. Brightness and loudness as functions of stimulus duration. Percept. and Psychophys., 1966, 1, 319-327.

Stevenson, D.C. and Stephens, R.C. A programming system for psychoacoustic experimentation. Proceedings of the 11th DECUS (Canada) Symposium, Ottawa, Feb. 15-17, 1978.

Studdert-Kennedy, M., Liberman, A.M., Harris, K.S. and Cooper, F.S. Motor theory of speech perception: A reply to Lane's critical review. Psych. Rev., 1970, 77, 234-249.

Studdert-Kennedy, M. and Shankweiler, D. Hemispheric specialization for speech perception. *J. Acoust. Soc. Am.*, 1970, *48*, 579-594.

Taylor, M.M. and Aldridge, K.D. Stochastic processes in reversing figure perception. *Percept. and Psychophys.*, 1974, *16*, 9-27.

Thom, R. *Structural Stability and Morphogenesis*. (Translated by D.H. Fowler) New York, Benjamin, 1975.

Trittipoe, W.J. Residual effects at longer pre-exposure durations. *J. Acoust. Soc. Am.*, 1958, *30*, 1017-1019.

Taylor, M.M. and Creelman, C.D. PEST: Efficient estimates on probability functions. *J. Acoust. Soc. Am.*, 1967, *41*, 782-787.

Ward, W.D., Glorig, A. and Sklar, D.L. Temporary threshold shift produced by intermittent exposure to noise. *J. Acoust. Soc. Am.*, 1959, *31*, 791-794.

Ward, W.D., Glorig, A. and Selters, W. Temporary threshold shift in a changing noise level. *J. Acoust. Soc. Am.*, 1960, *32*, 235-237.

Williams, L. The perception of stop consonant voicing by Spanish-English bilinguals. *Percept. and Psychophys.*, 1977, *21*, 289-297.

Wilson, H.R. and Cowan, J.D. Excitatory and inhibitory interactions in populations of model neurons. *Biophys. J.*, 1972, *12*, 1-24.

Wood, C.C. Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *J. Acoust. Soc. Am.*, 1976, *60*, 1381-1389.

Wright, H.N. Auditory adaptation in noise. *J. Acoust. Soc. Am.*, 1959, *31*, 1004-1012.

Yund, E.W. and Efron, R. Dichotic competition of simultaneous tone bursts of different frequency II. Suppression and ear dominance functions. *Neuropsychologia.*, 1975, *13*, 137-150.

Yund, E.W. and Efron, R. Dichotic competition of simultaneous tone bursts of different frequency IV. Correlation with dichotic competition of speech signals. *Brain and Lang.*, 1976, *3*, 246-254.

Yund, E.W. and Efron, R. Model for the relative salience of the pitch of pure tones presented dichotically. J. Acoust. Soc. Am., 1977, 62, 607-617.

Zinnes, J.L. and Kurtz, R. Matching, discrimination, and payoffs. Jour. Math. Psych., 1968, 5:392-421.

Zinnes, J.L. and Wolff, R.P. Single and multidimensional same-different judgements.

Zwislocki, J.J. Temporal summation of loudness: An analysis. J. Acoust. Soc. Am., 1969, 46, 431-441.

Zwislocki, J.J. On intensity characteristics of sensory receptors: A generalized function. Kybernetik, 1973, 12: 169-183')